On Some Questions of Input Signal Synthesis

Keviczky, Laszlo

1972

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Keviczky, L. (1972). *On Some Questions of Input Signal Synthesis*. (Research Reports TFRT-3047). Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

ON SOME QUESTIONS OF INPUT SIGNAL SYNTHESIS

Laszlo Keviczky

Report 7226 (B) november 1972
Lund Institute of Technology
Division of Automatic Control

ON SOME QUESTIONS OF INPUT SIGNAL SYNTHESIS

LASZLO KEVICZKY

ON SOME QUESTIONS OF INPUT SIGNAL SYNTHESIS

L.K. Keviczky

ABSTRACT

In this report some questions in connection with the design of input signals used for the parametric identification, i.e. some questions of "input signal synthesis", are considered. The problem is approached from the aspects of sensitivity and statistics. We have examined some possibilities to choose different criteria for the input synthesis. The basic differences between the static and dynamic cases are shown. We have established some input synthesis rules for the case of models having linearity in parameters. For discrete dynamic system models we have determined a control law which generates the locally optimum input by a real-time simulation and by an off-line method. The deduced relations are illustrated by simulation results.

TABLE OF CONTENTS

LIST OF SYMBOLS

$\underline{\underline{A}}$	–	matrix
$\underline{x}$	–	column-vector
$\underline{\underline{A}}^T$	–	transposed matrix
$\underline{x}^T$	–	row-vector, transposed column-vector
$\lvert\underline{\underline{A}}\rvert$	–	the determinant of matrix $\underline{\underline{A}}$
$tr(\underline{\underline{A}})$	–	the trace of matrix $\underline{\underline{A}}$
$E\{\ldots\}$	–	operator of mathematical expectation
z	–	the variable of $\mathcal{Z}$ transformation (or forward shift-operator)
z^{-1}	–	backward shift-operator
$\lVert\underline{x}\rVert=\sqrt{\underline{x}^T\underline{x}}$	–	the norm of vector $\underline{x}$
$\frac{d}{d\underline{x}}$	–	(column) vector differential operator
$\frac{d}{d\underline{x}^T}$	–	(row) vector differential operator
$\in$	–	element of a set
$\underline{\underline{I}}$	–	identity matrix

1. INTRODUCTION

It has been proved in practice that identification of processes, i.e. determination of structure and parameters has a great importance for the preparation and design of process control. The theory of identification has been approaching a closed and complete state though this can never be reached because of great variety of systems. The fact is that the functionals which provide the estimates of best behaviour have been established for structures having most importance in practice and the parameter estimation has been reduced to a mathematical programming (extremum seeking) procedure. This task can also be solved successfully for the multivariable, multiparameter systems when fast computers with large memories will be in more general use. The question of joint performing of identification and control has been coming to the focus of attention which constitutes the significant part of the process control by computer.

Similarly it can be of interest what kind of input signals provides the estimates of best behaviour for a given system under certain restrictions. The optimal input signal first of all depends on the chosen optimality criterion over the system equation and restrictions.

First in this report the sensitivity and statistical questions of input signal synthesis are considered. Then detailed examinations are made for static and discrete dynamic systems being linear in the parameters.

In this paper the methods and criteria of continuous design of experiments used for static model are briefly discussed. The experiences gained from the static case are used for optimum input synthesis of linear discrete-time systems. A simple algorithm is given by means of which the determinant of the information matrix can be maximized. The control rules of optimum input synthesis are investigated both in on-line and off-line case. These control rules generate locally optimum inputs because they optimize only the next step in every time point.

The cases of correlated and uncorrelated noise are both examined. Our theoretical computations and the importance of the optimal input signal synthesis have been verified by simulation results.

2.INPUT SIGNAL SYNTHESIS AND SENSITIVITY

To determine the optimum input we need a measure to get the input by its optimization. Let us examine this question from the point of view of sensitivity of output with respect to the parameter changing. Obviously two systems with the same structure and with different parameter values ,close to each other, can be distinguished the more easily the bigger the changing of output is in consequence of changing of parameters [16,17] .
Let us consider the output x of a system at the different values of parameters, $\underline{p}_1$ and $\underline{p}_2$, and let the difference of output values be Δx .Expanding the output in first order Taylor series

$$\Delta x = x(\underline{p}_2) - x(\underline{p}_1) \cong \left.\frac{dx(\underline{p})}{d\underline{p}^T}\right|_{\Delta\underline{p}=\underline{0}} \Delta\underline{p} \qquad (2.1)$$

where $\Delta\underline{p} = \underline{p}_2 - \underline{p}_1$ and the higher order terms are neglected. The easier to distinguish the two systems (i.e. the identification is easier) the bigger the so called output sensitivity(vector)function in (2.1) [17] .

$$\underline{s} = \left.\frac{dx(\underline{p})}{d\underline{p}}\right|_{\Delta\underline{p}=0} \qquad (2.2)$$

In general $\underline{s}$ depends on input signal u, output signal x, parameters p and the time t.

The optimal input which is necessary to the identification can be obtained by the maximization of this vector. Being the output sensitivity function vector so several kinds of measures can be formulated for its maximization.

Constituting the sensitivity matrix

$$\underline{\underline{S}} = \underline{s}\,\underline{s}^T \qquad (2.3)$$

we can choose the criterion as a measure

$$Q_1 = tr(\underline{\underline{S}}) = tr(\underline{s}\,\underline{s}^T) = \underline{s}^T\,\underline{s} = \|\underline{s}\|^2 \longrightarrow max \qquad (2.4)$$

i.e. the square of the norm of the sensitivity function; or

$$Q_2 = |\underline{\underline{S}}| \longrightarrow max \qquad (2.5)$$

i.e. the determinant of the sensitivity matrix;

$$Q_3 = \lambda_{min}(\underline{\underline{S}}) \longrightarrow max \qquad (2.6)$$

i.e. the maximization of smallest eigenvalue of the sensitivity matrix or for example

$$Q_4 = \sum_{i \neq j}^{k} \sum^{k} S_{ij}^2 \longrightarrow min \qquad (2.7)$$

i.e. a criterion for the diagonalization of sensitivity matrix and several types of criteria.

The extremum of one of these measures must be ensured by u(t) in every point t under some restrictions for u

$$u(t) \in \mathcal{U} \qquad (2.8)$$

where $\mathcal{U}$ is a set of the admissible input signals.

Let us formulate some definitions in connection with the input sugnal synthesis [17] .

Definition 1.

If for any input signal

$$\underline{s} = \frac{dx}{d\underline{p}} = \underline{const} \tag{2.9}$$

i.e. the output sensitivity function is independent of the input signal then the input is called unsynthetizable. If only one element of the sensitivity function is independent of input signal then the input is unsynthetizable only with respect to this parameter.

Definition 2.

If for any input signal

$$\underline{s} = \frac{dx}{d\underline{p}} = \underline{0} \tag{2.10}$$

then the system is called unidentifiable. In this case the sensitivity function disappears and the output of the two systems can not be distingiushed in first order approximation. (This is the condition for the slight invariance [17]).

If we want to determine an optimum input signal on the basis of criterion according to the sensitivity function such problems appear which are unsolvable or need much computation. The results obtained in this field are of theoretical importance but they have called the attention to the significance of input signal synthesis and to the similarity of this problems to the sensitivity ones.

Now let us turn to the optimal synthesis of input on the basis of statistics.

3. INPUT SIGNAL SYNTHESIS ON THE BASIS OF INFORMATION MATRIX

The disadvantage of sensitivity approach mentioned in the previous section is its deterministic character over and above the complicated computations. The simplest way to the stochastic approach is to take the mathematical expectation of measures (2.4) - (2.7) as a performance criterion.

As it will be shown we can get to the same result if we use the FISHER information matrix in our analysis. It is known that the lower bound of the accuracy of parameter estimation is given by the CRAMER-RAO inequality [3,19] .

$$E\left\{ (\hat{\underline{p}} - \underline{p})(\hat{\underline{p}} - \underline{p})^T \right\} = \underline{\underline{K}} \geq \underline{\underline{M}}^{-1} \qquad (3.1)$$

where $\hat{\underline{p}}$ is the estimated-, $\underline{p}$ is the original parameter vector, $\underline{\underline{K}}$ is their covariance matrix and $\underline{\underline{M}}$ is the so-called information matrix.

Let us assume that we have the values **y** of x disturbanced by the noise ε in a vector form for N samplings.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \underline{y} = \underline{x} + \underline{\varepsilon} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} \qquad (3.2)$$

According to the definition the information matrix is [12] :

$$\underline{\underline{M}} \triangleq \int L(\underline{y}|\underline{p}) \left[\frac{d}{d\underline{p}} \ln L(\underline{y}|\underline{p}) \right] \left[\frac{d}{d\underline{p}} \ln L(\underline{y}|\underline{p}) \right]^T d\underline{y} \qquad (3.3)$$

where $L(\underline{y}|\underline{p})$ is conditional probability density function of $\underline{y}$. Assuming that the measuring noise $\underline{\varepsilon}$ is of normal distribution and

$$E\{\underline{\varepsilon}\} = \underline{0} \qquad E\{\underline{\varepsilon}\,\underline{\varepsilon}^T\} = \underline{\underline{W}}^{-1} \tag{3.4}$$

then on the basis of computations in Appendix A

$$\underline{\underline{M}} = \frac{d\underline{x}^T}{d\underline{p}}\,\underline{\underline{W}}\,\frac{d\underline{x}}{d\underline{p}^T} \tag{3.5}$$

In a special case when the measuring noise is uncorrelated

$$E\{\underline{\varepsilon}\,\underline{\varepsilon}^T\} = \underline{\underline{W}}^{-1} = \lambda^2\,\underline{\underline{I}} \tag{3.6}$$

and the information matrix

$$\underline{\underline{M}} = \underline{\underline{M}}_N = \frac{1}{\lambda^2}\,\frac{d\underline{x}^T}{d\underline{p}}\,\frac{d\underline{x}}{d\underline{p}^T} = \frac{1}{\lambda^2}\sum_{i=1}^{N}\frac{dx_i}{d\underline{p}}\,\frac{dx_i}{d\underline{p}^T} \tag{3.7}$$

On the basis of this we have

$$\lambda^2\,\underline{\underline{M}}_N = \lambda^2\,\underline{\underline{M}}_{N-1} + \frac{dx_N}{d\underline{p}}\,\frac{dx_N}{d\underline{p}^T} \tag{3.8}$$

Comparing (2.3) with (3.7) it gives

$$E\{\underline{\underline{S}}\} = E\{\underline{s}\,\underline{s}^T\} = \lim_{N\to\infty}\frac{\lambda^2}{N}\,\underline{\underline{M}} \tag{3.9}$$

i.e. in the case of uncorrelated noise the information matrix is proportional to the expectation value of the sensitivity matrix for $N\to\infty$. If we want to produce optimal input signal from the stochastical point of view, then the criteria of optimality must be formulated for the information matrix. These may be of types according to

(2.4) - (2.7), but several others, too.

Let us denote $\lambda^2 \underline{\underline{M}}$ with $\underline{\underline{G}}$, so

$$\underline{\underline{G}}_N = \underline{\underline{G}}_{N-1} + \underline{s}_N \underline{s}_N^T \tag{3.10}$$

Since $\underline{\underline{M}}^{-1}$ is needed to the CRAMER-RAO lower bound, so let us form $\underline{\underline{G}}^{-1}$. It can be obtained from (3.10) by the well-known inversion theory of matrices that

$$\underline{\underline{G}}_N^{-1} = \underline{\underline{G}}_{N-1}^{-1} - \frac{\underline{\underline{G}}_{N-1}^{-1} \underline{s}_N \underline{s}_N^T \underline{\underline{G}}_{N-1}^{-1}}{1 + \underline{s}_N^T \underline{\underline{G}}_{N-1}^{-1} \underline{s}_N} \tag{3.11}$$

or

$$\underline{\underline{M}}_N^{-1} = \underline{\underline{M}}_{N-1}^{-1} - \frac{\underline{\underline{M}}_{N-1}^{-1} \underline{s}_N \underline{s}_N^T \underline{\underline{M}}_{N-1}^{-1}}{\lambda^2 + \underline{s}_N^T \underline{\underline{M}}_{N-1}^{-1} \underline{s}_N} \tag{3.12}$$

In the case of correlated noise performing the factorization

$$\underline{\underline{W}} = \frac{1}{\lambda^2} \underline{\underline{T}}^T \underline{\underline{T}} \tag{3.13}$$

formally we can get to the (3.7) again by the transformation (filtering)

$$\underline{x}^F = \underline{\underline{T}} \, \underline{x} \quad , \tag{3.14}$$

i.e.

$$\underline{\underline{M}} = \frac{1}{\lambda^2} \quad \frac{d\underline{x}^{F^T}}{d\underline{p}} \frac{d\underline{x}^F}{d\underline{p}^T} \tag{3.15}$$

Remark:

now the relation in connection with the sensitivity function is not so striking since in the case of correlated noise

$$E\left\{\underline{\underline{S}}^F\right\} = E\left\{\underline{s}^F \underline{s}^{F^T}\right\} = \lim_{N \to \infty} \frac{\lambda^2}{N} \underline{\underline{M}} \tag{3.16}$$

where $\underline{s}^F$ is the sensitivity function of fictive output formed by the filtering (3.14).

Further on some simple and utilizable methods will be given for the optimal input signal synthesis.The treatment is limited to the systems being linear in parameters. In the case of systems being nonlinear in parameters there are several serious computational problems as it was mentioned above (See [14]) .

Furthermore we are going to deal with such strategies which maximize the determinant of the information matrix. This strategy corresponds to the fact that the optimal input signal maximizes the information about the process [11].

Turning from an estimation procedure elaborating N samplings to an other one elaborating N+1 samplings the increasing rate of information is defined by STONE [9] (on the basis of BAYES' work):

$$\Delta I = I\left\{L(\hat{\underline{p}}|\underline{y}_{N+1})\right\} - I\left\{L(\hat{\underline{p}}|\underline{y}_N)\right\} \tag{3.17}$$

where

$$I\left\{L(\underline{v})\right\} = \int L(\underline{v}) \log L(\underline{v})\, d\underline{v} \tag{3.18}$$

Here $L(\hat{\underline{p}}|\underline{y}_N)$ is an apriori , $L(\hat{\underline{p}}|\underline{y}_{N+1})$ is a posteriori probability density function. Assuming normal distribution it can be obtained (see in Appendix B) that

$$\Delta I = \frac{1}{2} \log \frac{|\underline{\underline{K}}_{N+1}^{-1}|}{|\underline{\underline{K}}_N^{-1}|} \tag{3.19}$$

So in the case of input signal synthesis that is the optimal strategy locally which maximizes the increase of the determinant of inverse of covariance matrix in every step. In static case (see in Section 4.) when the LS estimation is an efficient estimation, $\underline{\underline{K}}_N^{-1} = \underline{\underline{M}}_N$ and the former optimal input maximizes the determinant of the information matrix in every step and minimizes the volume of ellipsoid according to the CRAMER-RAO lower bound.

In discrete dynamic model case (see in Section 5), when the estimation is asymptotically efficient there are two possibilities. One of them is to maximize the determinant of the inverse of covariance matrix, i.e. information about the process, the other is to maximize the determinant of information matrix which is optimal for $N \to \infty$.

4. INPUT SIGNAL SYNTHESIS FOR STATIC MODEL

For the static model identification the optimal input synthesis is called design of experiments.[8]

Keeping the condition mentioned above in connection with the linearity in the parameters let the static characteristic of the process be

$$y = \underline{f}^T(\underline{u})\ \underline{p} + \varepsilon = x + \varepsilon \qquad (4.1)$$

(See Fig.1.)

where

$$\underline{u} = [u_1, \ldots, u_n]^T \qquad (4.2)$$

is an n vector of input variables,

$$\underline{f}(\underline{u}) = [f_1(\underline{u}), \ldots, f_k(\underline{u})]^T \qquad (4.3)$$

is a k vector of function components and $\underline{p}$ is a k vector of the true parameter values. It is assumed that ε is uncorrelated, independent of $\underline{u}$ and has zero mean value and constant standard deviation.

If we have N sampled values of $\underline{u}$ and y then the well-known LS estimation of parameters is:

$$\hat{\underline{p}} = (\underline{\underline{F}}^T\ \underline{\underline{F}})^{-1}\ \underline{\underline{F}}^T\ \underline{y} \qquad (4.4)$$

Here

$$\underline{\underline{F}} = \begin{bmatrix} \underline{f}^T(\underline{u}_1) \\ \vdots \\ \underline{f}^T(\underline{u}_N) \end{bmatrix} \qquad (4.5)$$

and

$$\underline{y} = [y_1, \ldots, y_N]^T \qquad (4.6)$$

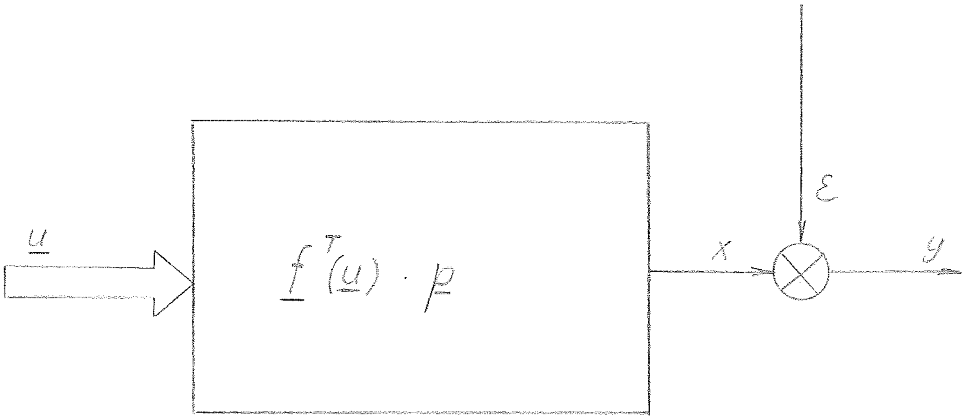


Fig. 1

Moreover, the covariance matrix of estimation (4.4)

$$\underline{\underline{K}} = E\left\{(\hat{\underline{p}} - \underline{p})(\hat{\underline{p}} - \underline{p})^T\right\} = (\underline{\underline{F}}^T \underline{\underline{F}})^{-1} \lambda^2 \tag{4.7}$$

Since in this case

$$\underline{s} = \frac{dx}{d\underline{p}} = \underline{f}(\underline{u}) \tag{4.8}$$

so on the basis of (3.7) the information matrix is:

$$\underline{\underline{M}} = \frac{1}{\lambda^2} \sum_{i=1}^{N} \frac{dx_i}{d\underline{p}} \frac{dx_i}{d\underline{p}^T} = \frac{1}{\lambda^2} \sum_{i=1}^{N} \underline{f}(\underline{u}_i)\underline{f}^T(\underline{u}_i) = \frac{1}{\lambda^2} (\underline{\underline{F}}^T \underline{\underline{F}}) \tag{4.9}$$

Thus it is valid that

$$\underline{\underline{K}} = (\underline{\underline{F}}^T \underline{\underline{F}})^{-1} \lambda^2 = \underline{\underline{M}}^{-1} \tag{4.10}$$

i.e. the covariance matrix of the parameter estimates coincides with the CRAMER-RAO lower bound so (4.4) is an efficient estimation for every N. Thus the optimal input synthesis requires the maximization of $\underline{\underline{M}}$, i.e. the minimization of $\underline{\underline{K}}$. In the case of LS estimation the loss function

$$V = \frac{1}{2}(\underline{y} - \underline{\underline{F}}\,\hat{\underline{p}})^T(\underline{y} - \underline{\underline{F}}\,\hat{\underline{p}}) \tag{4.11}$$

is minimized for which the HESSIAN-matrix of the second order derivatives is:

$$\underline{\underline{H}}(V,\hat{\underline{p}}) = \frac{d^2V}{d\hat{\underline{p}}\; d\hat{\underline{p}}^T} = \underline{\underline{F}}^T\underline{\underline{F}} \tag{4.12}$$

In according to this the maximization of $\underline{\underline{M}}$ also maximizes the curvature of the loss function in the environment of extremum.

A prediction can be made by the estimated value of $\underline{p}$

$$\hat{y}(\underline{u}) = \underline{f}^T(\underline{u})\,\hat{\underline{p}} \tag{4.13}$$

and its variance is:

$$\mathrm{var}\{\hat{y}\} = E\{[\hat{y}(\underline{u}) - E\{\hat{y}(\underline{u})\}]^2\} = \underline{f}^T(\underline{u})\ \underline{\underline{K}}\ \underline{f}(\underline{u}) =$$
$$= \underline{f}^T(\underline{u})(\underline{\underline{F}}^T\ \underline{\underline{F}})^{-1}\ \underline{f}(\underline{u})\ \lambda^2 \qquad (4.14)$$

In the theory of experimental design the following optimality criteria respect to the information matrix (to the covariance matrix, respectively) are used [8] :

1. Orthogonality: i.e. let the covariance matrix be diagonal. This requirement makes it possible that the parameters can be estimated statistically independently from each other.
2. Rototabellity: This criterion demands that $\underline{\underline{K}}$ must be invariant to the orthogonal transformation of variables. This means that the level contours of predicted variance must be rotosymmetrical (i.e. spherical) in the space of input variables.
3. A-optimality: This requires that the trace of covariance matrix i.e. the mean value of variances of parameters must be minimum.
4. E-optimality: According to this criterion the largest eigenvalue of the covariance matrix must be minimized.
5. G-optimality: According to this the largest value of the variance of prediction is minimized on a given constrained space.
6. D-optimality: In this case the determinant of the covariance matrix is minimized (or the determinant of the information matrix is maximized).

In addition to those criteria several other criteria can be constructed, only the most important ones have been mentioned above.

An important thing in connection with their application is that their optimization has meaning only for a given constrained space of the independent variables (otherwise $\|\underline{u}\| \to \infty$ is a trivial optimal solution).

Such a series $\underline{u}_1, \ldots, \underline{u}_N$ of the independent variables for which a certain optimality criterion is valid

$$Q(\underline{\underline{M}}) = \underset{\underline{u}_1, \ldots, \underline{u}_N}{\text{extremum}} \quad , \quad \underline{u}_i \in \mathcal{U} \tag{4.15}$$

is called optimum experimental design.

There are two tasks at choosing of the optimum experimental design: in one case the number N of the combinations of input variables is given, in other case the criterion must be optimized for $N \to \infty$.

In the first case finite, in the second case infinite or continuous design is considered.

Since according to (4.9) $\underline{\underline{M}}$ depends only on the form of function $\underline{f}(\underline{u})$ and measuring points $\underline{u}_i$ and is independent of y so the basic theory of design of experiments can be formulated [20].

Theory: In the case of process model being linear in parameters if the observation error is an uncorrelated, independent noise of the input signal with zero mean value and identical variance, then the optimum input series (experimental design) according to the criterion with respect to information matrix can be determined in advance before the measuring independently of the process.

Thus the optimal experimental design depends only on the structure of static characteristic and the form of the admissible constrained space for the input signals.
In general the synthesis of finite designs is very difficult problem [8] now we do not intend to deal with them - and from the continuous ones only with the synthesis of A- and D-optimal design - because we are going to use these methods in the case of dynamic process model, too.
Using the definition (3.10) of $\underline{\underline{G}}$ it can be written for the static case

$$\underline{\underline{G}}_{N+1} = \underline{\underline{G}}_N + \underline{f}(\underline{u}_{N+1})\ \underline{f}^T(\underline{u}_{N+1}) \tag{4.16}$$

Then it can be proved [19] that the following identity exists:

$$\frac{|\underline{\underline{G}}_{N+1}|}{|\underline{\underline{G}}_N|} = 1 + \underline{f}^T(\underline{u}_{N+1})\ \underline{\underline{G}}_N^{-1}\ \underline{f}(\underline{u}_{N+1}) \tag{4.17}$$

or in another way

$$\frac{|\underline{\underline{M}}_{N+1}|}{|\underline{\underline{M}}_N|} = 1 + \frac{1}{\lambda^2}\ \underline{f}^T(\underline{u}_{N+1})\ \underline{\underline{M}}_N^{-1}\underline{f}(\underline{u}_{N+1}) = 1 + \frac{1}{\lambda^2}\ \underline{f}^T(\underline{u}_{N+1})\underline{\underline{K}}_N\underline{f}(\underline{u}_{N+1}) = 1 + \frac{\mathrm{var}\{\hat{y}(\underline{u}_{N+1})\}}{\lambda^2} \tag{4.18}$$

From the different presentations the following important statements can be drawn. The locally optimal strategy maximizing the determinant of the information matrix is that in which the next optimal input $\underline{u}_{N+1}^o$ is chosen on such way that

$$\frac{1}{\lambda^2}\ \underline{f}^T(\underline{u}_{N+1}^o)\ \underline{\underline{K}}_N\underline{f}(\underline{u}_{N+1}^o) = \max_{\underline{u}\in\mathcal{U}} \underline{f}^T(\underline{u})\ \underline{\underline{G}}_N^{-1}\ \underline{f}(\underline{u}) =$$

$$= \max_{\underline{u} \in \mathcal{U}} \frac{\operatorname{var}\{\hat{y}(\underline{u})\}}{\lambda^2} \tag{4.19}$$

where $\mathcal{U}$ is the admissible set of input variable.

Let us take into account that

$$\underline{\underline{G}}_{N+1}^{-1} = \underline{\underline{G}}_N^{-1} - \frac{\underline{\underline{G}}_N^{-1}\ \underline{f}(\underline{u}_{N+1})\ \underline{f}^T(\underline{u}_{N+1})\ \underline{\underline{G}}_N^{-1}}{1 + \underline{f}^T(\underline{u}_{N+1})\ \underline{\underline{G}}_N^{-1}\ \underline{f}(\underline{u}_{N+1})} \tag{4.20}$$

Multiplying this equality by $\underline{f}(\underline{u}_{N+1})$ from the right side and by $\underline{f}^T(\underline{u}_{N+1})$ from the left side we get that

$$\underline{f}^T(\underline{u}_{N+1})\ \underline{\underline{G}}_{N+1}^{-1}\ \underline{f}(\underline{u}_{N+1}) = \frac{\underline{f}^T(\underline{u}_{N+1})\ \underline{\underline{G}}_N^{-1}\ \underline{f}(\underline{u}_{N+1})}{1+\underline{f}^T(\underline{u}_{N+1})\ \underline{\underline{G}}_N^{-1}\underline{f}(\underline{u}_{N+1})} \tag{4.21}$$

In according to this

$$\operatorname{var}\{\hat{y}(\underline{u}_{N+1})|\underline{y}_{N+1}\} = \frac{1}{1+\underline{f}^T(\underline{u}_{N+1})\ \underline{\underline{G}}_N^{-1}\underline{f}(\underline{u}_{N+1})} \operatorname{var}\{\hat{y}(\underline{u}_{N+1})|\underline{y}_N\} \tag{4.22}$$

This latter expression means that the continuous D and G-optimality coincide for static model,since this strategy chooses the next point so that the variance of prediction is maximum [8] .

Thus a continuous D-optimal design can be created in the following way. Starting from a non-singular matrix $\underline{\underline{G}}_o$ and solving the extremum problem under restrictions according to the (4.19) we get the first value $\underline{u}_1$ then update $\underline{\underline{G}}$ according to (4.20) and continue the procedure sequentially. (It can be proved that when choosing $\underline{\underline{G}}_o$ only the non-singularity is important [10].)

This designing method increases the determinant of the information matrix in optimum (i.e. maximum) rate in every step.

Let us examine how to choose the input signal for A-optimal design. The trace of the covariance matrix of the parameter estimates is

$$\mathrm{tr}(\underline{\underline{K}}) = \chi^2 \, \mathrm{tr}(\underline{\underline{G}}^{-1}) \tag{4.23}$$

On the basis of (4.20)

$$\mathrm{tr}(\underline{\underline{G}}_{N+1}^{-1}) = \mathrm{tr}(\underline{\underline{G}}_{N}^{-1}) - \mathrm{tr} \frac{\underline{\underline{G}}_{N}^{-1} \, \underline{f}(\underline{u}_{N+1}) \, \underline{f}^{T}(\underline{u}_{N+1}) \, \underline{\underline{G}}_{N}^{-1}}{1 + \underline{f}^{T}(\underline{u}_{N+1}) \, \underline{\underline{G}}_{N}^{-1} \, \underline{f}(\underline{u}_{N+1})} =$$

$$= \mathrm{tr}(\underline{\underline{G}}_{N}^{-1}) - \frac{\underline{f}^{T}(\underline{u}_{N+1}) \, \underline{\underline{G}}_{N}^{-1} \, \underline{\underline{G}}_{N}^{-1} \, \underline{f}(\underline{u}_{N+1})}{1 + \underline{f}^{T}(\underline{u}_{N+1}) \, \underline{\underline{G}}_{N}^{-1} \, \underline{f}(\underline{u}_{N+1})} \tag{4.24}$$

Thus the locally optimal strategy minimizing the trace of the covariance matrix is that in which the next optimum $\underline{u}_{N+1}^{o}$ is chosen on such way that

$$\frac{\underline{f}^{T}(\underline{u}_{N+1}^{o}) \, \underline{\underline{G}}_{N}^{-1} \, \underline{\underline{G}}_{N}^{-1} \, \underline{f}(\underline{u}_{N+1}^{o})}{1 + \underline{f}^{T}(\underline{u}_{N+1}^{o}) \, \underline{\underline{G}}_{N}^{-1} \, \underline{f}(\underline{u}_{N+1}^{o})} = \max_{\underline{u} \in \mathcal{U}} \tag{4.25}$$

must be valid [20].

5. INPUT SIGNAL SYNTHESIS FOR LINEAR DISCRETE-TIME DYNAMIC SYSTEMS

It has been mentioned that the input signal synthesis can be solved relatively easily for static models being linear in parameters and the task has been reduced to a multivariable extremum seeking problem with restrictions. In the case of static identification wide class of functions belongs to these types of characteristics. The dynamic identification is non-linear estimation problem almost in every case. LEVADI has tried to find algorithms by expanding linear power series of output in order to determine the optimal input signal but he encountered serious computational problems [14] .
AOKI and STALEY [12] have called attention to the fact that for identification model used at the linear discrete-time dynamic systems the quasilinearization ("multiplying by the denominator") can be performed easily enough but an additional effect is that it also changes the noise structure. In this section the synthesis of optimal input of linear,discrete-time,dynamic systems is considered.
In our investigations the pulse transfer function is used as a process model in the following form:

$$A(z^{-1})\, y(t) = B(z^{-1})\, u(t) + \lambda C(z^{-1})\, e(t) \qquad (5.1)$$

$$t = 1,2,\ldots,N$$

where

$\{y(t),\ t=1,2,\ldots,N\}$ – is the observed output signal

$\{u(t), t=1,1,\ldots,N\}$ – is the applied input signal

$\{e(t), t=1,2,\ldots,N\}$ – is a disturbance sequence of independent normal random variables of zero mean and variance one.

z^{-1} – is the backward shift-operator.

$$A(z^{-1}) = 1+a_1z^{-1} + \ldots + a_nz^{-n} = 1 + \tilde{A}(z^{-1})$$
$$B(z^{-1}) = (b_o) + b_1z^{-1} + \ldots + b_mz^{-m}$$
$$C(z^{-1}) = 1 + c_1z^{-1} + \ldots + c_nz^{-n}$$

n is the order of the model

N is the number of samplings

(See Fig.2.)

If we settle the values y(t), u(t), e(t) to a vector $\underline{y}$, $\underline{u}$, $\underline{e}$, respectively,with respect to t = 1,...,N then the relation is in vector form:

$$\underline{\underline{A}}\ \underline{y} = \underline{\underline{B}}\ \underline{u} + \lambda \underline{\underline{C}}\ \underline{e} \tag{5.2}$$

where

$$\underline{\underline{A}} = \underline{\underline{I}} + \sum_{i=1}^{n} a_i \underline{\underline{S}}^i = \underline{\underline{I}} + \tilde{\underline{\underline{A}}} \tag{5.3}$$

$$\underline{\underline{B}} = (b_o \underline{\underline{I}}) + \sum_{i=1}^{m} b_i \underline{\underline{S}}^i \tag{5.4}$$

$$\underline{\underline{C}} = \underline{\underline{I}} + \sum_{i=1}^{n} c_i \underline{\underline{S}}^i \tag{5.5}$$

are so-called generalized TOEPLITZ matrices and $\underline{\underline{S}}^i$ is a shift-matrix which has ones in every i^{th} row under the diagonal elements,

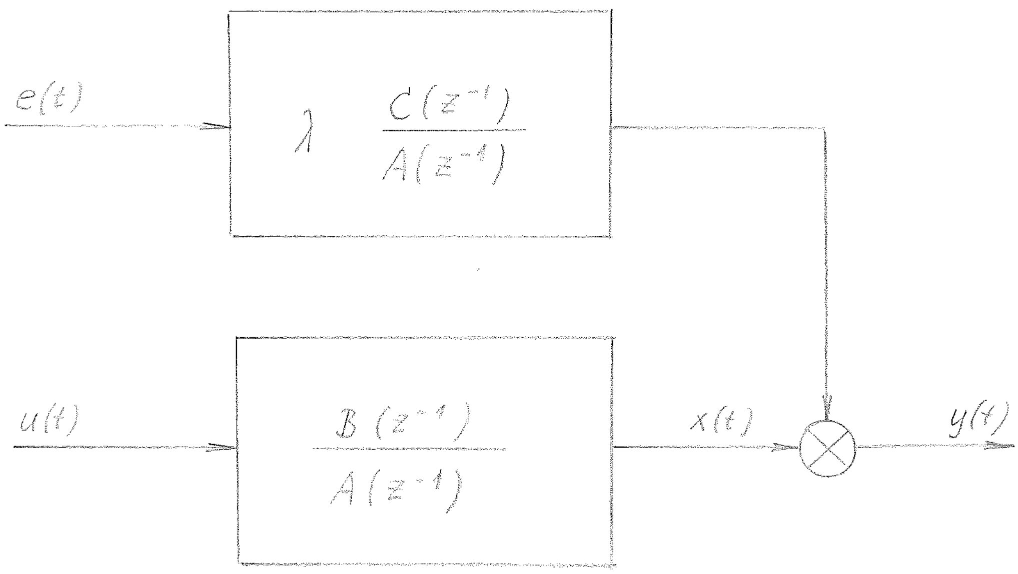


Fig. 2.

otherwise zeroes [18] .

Rearranging (5.2) we get

$$\underline{y} = \underline{\underline{B}}\ \underline{u} - \underline{\underline{\tilde{A}}}\ \underline{y} + \lambda\, \underline{\underline{C}}\ \underline{e} = \underline{\underline{F}}(\underline{u},\underline{y})\ \underline{p} + \lambda\, \underline{\underline{C}}\ \underline{e} \tag{5.6}$$

where $\underline{p}$ is a $k=m+n(+1)$ vector of process parameters

$$\underline{p} = \left[(b_o), b_1, \ldots, b_m, a_1, \ldots, a_n \right]^T \tag{5.7}$$

and

$\underline{\underline{F}}(\underline{u},\underline{y})$ is a HENKEL-matrix

$$\underline{\underline{F}}(\underline{u},\underline{y}) = \left[(\underline{\underline{S}}^o\ \underline{u}), \underline{\underline{S}}\ \underline{u}, \ldots, \underline{\underline{S}}^m\ \underline{u}, -\underline{\underline{S}}\underline{y}, \ldots, -\underline{\underline{S}}^n\ \underline{y} \right] = \underline{\underline{F}}_{uy} \tag{5.8}$$

The following scalar form being linear in parameters corresponds to the equation (5.6)

$$\begin{aligned} y(t) &= B(z^{-1})\ u(t) - \tilde{A}(z^{-1})\ y(t) + \lambda C(z^{-1})\ e(t) = \\ &= \underline{f}^T(u,y,t)\ \underline{p} + \lambda C(z^{-1})\ e(t) \end{aligned} \tag{5.9}$$

where

$$\underline{f}^T(u,y,t) = \left[(u(t)), u(t-1), \ldots, u(t-m), -y(t-1), \ldots, -y(t-n) \right] \tag{5.10}$$

Since the vector $\underline{f}(u,y,t)$ of function components depends not only on the input but also the output, the covariance matrix or information matrix of the estimates of different types in this case depend on the parameters of system. This means that the input signal which provides the required optimal property of covariance matrix (or information matrix) can be synthetised only with knowledge of the parameters. This can be approached by two methods. In the first case some results can be obtained by the recursive estimation of process parameters and the real-time simulation of the process (this will be examined in the case of LS structure), the other possibility is to use the parameter estimation obtained by an off-line identification to determine the optimum input series to the next off-line estimation. (This will be considered for the model (5.1) .)

5.1. CASE OF LS STRUCTURE

First let us consider the case when the equation error is white noise (uncorrelated) in (5.9), i.e. when $C(z^{-1}) \equiv 1$. The system is on Fig.3. In this case the well-known LS estimation [5] is unbiased

$$\hat{\underline{p}} = (\underline{\underline{F}}_{uy}^T \underline{\underline{F}}_{uy})^{-1} \underline{\underline{F}}_{uy}^T \underline{y} \tag{5.1.1}$$

and the covariance matrix of the estimation is

$$\underline{\underline{K}} = E\{(\hat{\underline{p}} - \underline{p})(\hat{\underline{p}} - \underline{p})^T\} = (\underline{\underline{F}}_{uy}^T \underline{\underline{F}}_{uy})^{-1} \lambda^2 = \left[\sum_{t=1}^{N} \underline{f}(u,y,t) \underline{f}^T(u,y,t)\right]^{-1} \lambda^2 \tag{5.1.2}$$

assuming that the number of samplings is N.

The recursive form of the LS estimation is:

$$\hat{\underline{p}}_{N+1} = \hat{\underline{p}}_N + \underline{\underline{R}}[N+1][y(N+1) - \underline{f}^T(u,y,N+1) \hat{\underline{p}}_N] \underline{f}(u,y,N+1) \tag{5.1.3}$$

where

$$\underline{\underline{R}}[N+1] = \underline{\underline{R}}[N] - \frac{\underline{\underline{R}}[N] \underline{f}(u,y,N+1) \underline{f}^T(u,y,N+1) \underline{\underline{R}}[N]}{1 + \underline{f}^T(u,y,N+1) \underline{\underline{R}}[N] \underline{f}(u,y,N+1)} = \frac{1}{\lambda^2} \underline{\underline{K}}_{N+1} \tag{5.1.4}$$

Now we are going to examine how to choose u(N+1) in order to fulfil different optimality criteria while the data are processed according to (5.13) and (5.14). Assume that the following amplitude restriction is valid for the input

$$-U \leq u(t) \leq +U \tag{5.1.5}$$

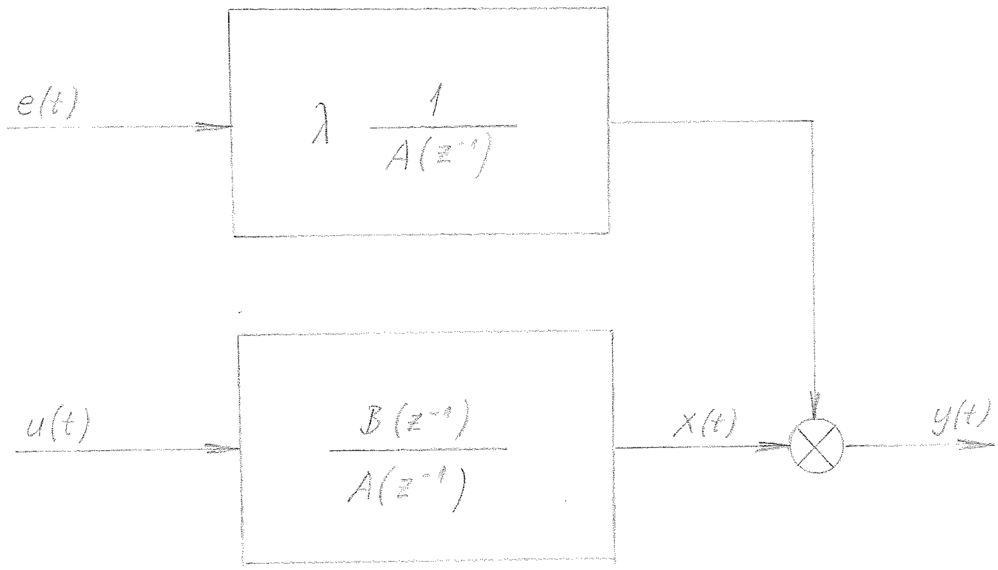


Fig. 3.

Maximization of $tr(\underline{\underline{K}}_N^{-1})$

From the point of view of the maximization of $tr(\underline{\underline{K}}_N^{-1})$ and further investigations it is reasonable to partitionate $\underline{f}(u,y,t)$ in the following manner:

$$\underline{f}(u,y,t) = [u(t),u(t-1),\dots,u(t-m),-y(t-1),\dots,-y(t-n)]^T = \\ = [u(t), \underline{g}^T(t-1)]^T \qquad (5.1.6)$$

Since according to (5.14) $\underline{\underline{K}}$ is proportional to $\underline{\underline{R}}$ we can examine only $\underline{\underline{R}}$. It can be written

$$\underline{\underline{R}}^{-1}[N+1] = \underline{\underline{R}}^{-1}[N] + \underline{f}(u,y,N+1)\,\underline{f}^T(u,y,N+1) \qquad (5.1.7)$$

Hence

$$tr(\underline{\underline{R}}^{-1}[N+1]) = tr(\underline{\underline{R}}^{-1}[N]) + \underline{f}^T(u,y,N+1)\,\underline{f}(u,y,N+1) = \\ = tr(\underline{\underline{R}}^{-1}[N]) + \Delta[N+1] \qquad (5.1.8)$$

Thus

$$\Delta[N+1] = [u(N+1), \underline{g}^T(N)][u(N+1), \underline{g}^T(N)]^T = \\ = u^2(N+1) + \underline{g}^T(N)\,\underline{g}(N) \qquad (5.1.9)$$

According to this it can be said only that the trace of the inverse of covariance matrix can be maximized if $u(t)$ is on the bound point under restrictions (5.1.5) in every step.

Maximization of $|\underline{\underline{K}}_N^{-1}|$

Comparing (4.16),(4.17) with (5.1.4),(5.1.7), we get

$$\frac{|\underline{\underline{K}}_{N+1}^{-1}|}{|\underline{\underline{K}}_{N}^{-1}|} = 1 + \underline{f}^{T}(u,y,N+1)\ \underline{\underline{R}}[N]\ \underline{f}(u,y,N+1) \qquad (5.1.10)$$

Thus a locally optimal strategy can be formed for the maximization of $|\underline{\underline{K}}^{-1}|$ if $u(t)$ maximizes the quadratic form on the right side of (5.1.10) in every step.

Let us partitionate $\underline{\underline{R}}[N]$ according to (5.1.6) in the following way:

$$\underline{\underline{R}}[N] = \left[\begin{array}{c|c} r_N & \underline{d}_N^{T} \\ \hline \underline{d}_N & \underline{\underline{Q}}_N \end{array}\right] = \underline{\underline{K}}_N \qquad (5.1.11)$$

By these notations we get for (5.1.10), that

$$\frac{|\underline{\underline{K}}_{N+1}^{-1}|}{|\underline{\underline{K}}_{N}^{-1}|} = u^{2}(N+1)\ r_N + 2\ u(N+1)\ \underline{g}^{T}(N)\ \underline{d}_N + \underline{g}^{T}(N)\ \underline{\underline{Q}}_N\ \underline{g}(N) + 1 \qquad (5.1.12)$$

i.e. this is a parabola as a function of $u(N+1)$ having its peak down whose minimum is in the point

$$u^{*}(N+1) = -\ \frac{\underline{g}^{T}(N)\ \underline{d}_N}{r_N} \qquad (5.1.13)$$

(Here r_N is positive in consequence of its meaning.)

In order to determine the optimal input the global maximum of this parabola is to be sought on a domain given by (5.1.5). This can be made according to a simple rule is presented on Fig.4. On the basis of Fig.4. - which gives the global optimum at $u^{o}(N+1)$ - $u^{o}(N+1)$ can be obtained according to the following formula:

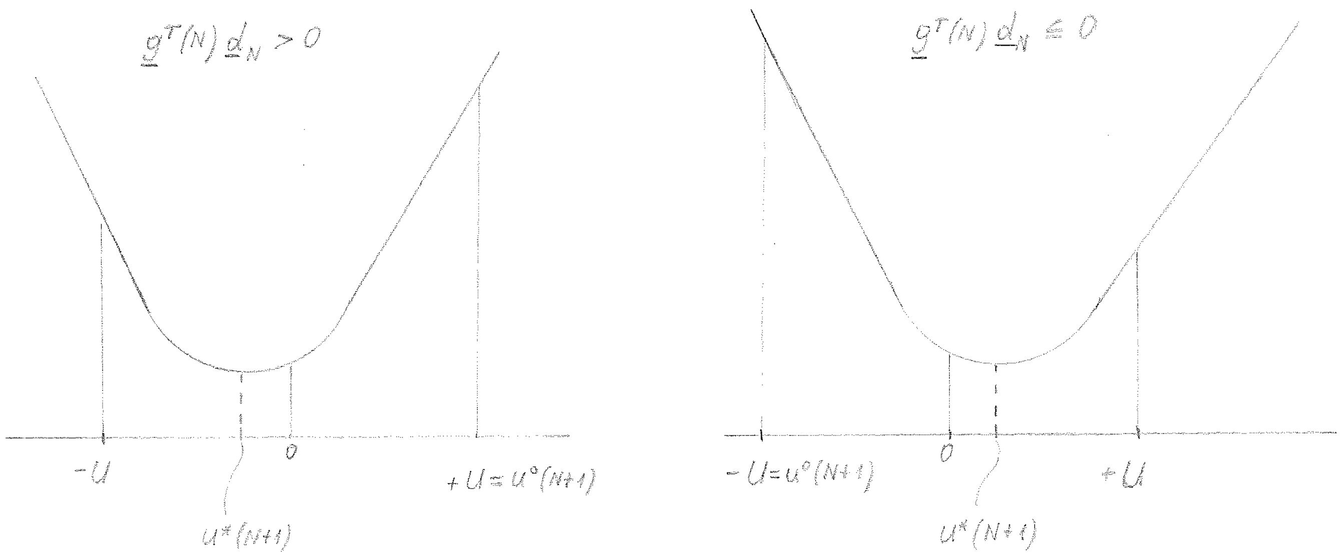

$\underline{g}^T(N)\,\underline{d}_N > 0$
$-U$
0
$+U = u^o(N+1)$
$u^*(N+1)$
$\underline{g}^T(N)\,\underline{d}_N \leq 0$
$-U = u^o(N+1)$
0
$+U$
$u^*(N+1)$

Fig. 4.

$$u^{o}(N+1) = \begin{cases} +U & \text{if } -\dfrac{\underline{g}^{T}(N)\,\underline{d}_N}{r_N} < 0 \quad \text{or} \quad \underline{g}^{T}(N)\,\underline{d}_N > 0 \\ -U & \text{if } -\dfrac{\underline{g}^{T}(N)\,\underline{d}_N}{r_N} \geq 0 \quad \text{or} \quad \underline{g}^{T}(N)\,\underline{d}_N \leq 0 \end{cases} \tag{5.1.14}$$

It can be seen that $u^{*}(t)$ and by this way $u^{o}(t)$, i.e. $u(t)$ depends only on $y(t-1),\dots,y(t-n)$,thus only on the $e(t-1),\dots,e(t-n)$. This means that the independence of the input and measuring noise is valid in this case,too.

On the basis of the deduced rule the optimum input signal synthesis can be performed according to the Fig.5.,where the synthesis of input and the parameter estimation are made according to the Fig.6. and Eqs.(5.1.3),(5.1.4),respectively.

Minimization of $tr(\underline{\underline{K}}_N)$

Since $\underline{\underline{K}}$ is proportional to $\underline{\underline{R}}$ now only $\underline{\underline{R}}$ is used. Comparing (4.20), (4.24) with (5.1.4) we get

$$tr(\underline{\underline{R}}[N+1]) = tr(\underline{\underline{R}}[N]) - \frac{\underline{f}^{T}(u,y,N+1)\,\underline{\underline{R}}[N]\,\underline{\underline{R}}[N]\,\underline{f}(u,y,N+1)}{1 + \underline{f}^{T}(u,y,N+1)\,\underline{\underline{R}}[N]\,\underline{f}(u,y,N+1)} \tag{5.1.15}$$

Taking into account (5.1.6) and (5.1.11) the second term of the right side of (5.1.15) can be written as follows:

$$\frac{(r_N^2 + \underline{d}_N^T\,\underline{d}_N)\,u^2(N+1) + 2(r_N\underline{d}_N^T + \underline{d}_N^T\underline{\underline{Q}}_N)\underline{g}(N)u(N+1) + \underline{g}^T(N)(\underline{d}_N\underline{d}_N^T + \underline{\underline{Q}}_N^2)\,\underline{g}(N)}{r_N^2\,u^2(N+1) + 2\,\underline{d}_N^T\,\underline{g}(N)\,u(N+1) + \underline{g}^T(N)\,\underline{\underline{Q}}_N\,\underline{g}(N) + 1} \tag{5.1.16}$$

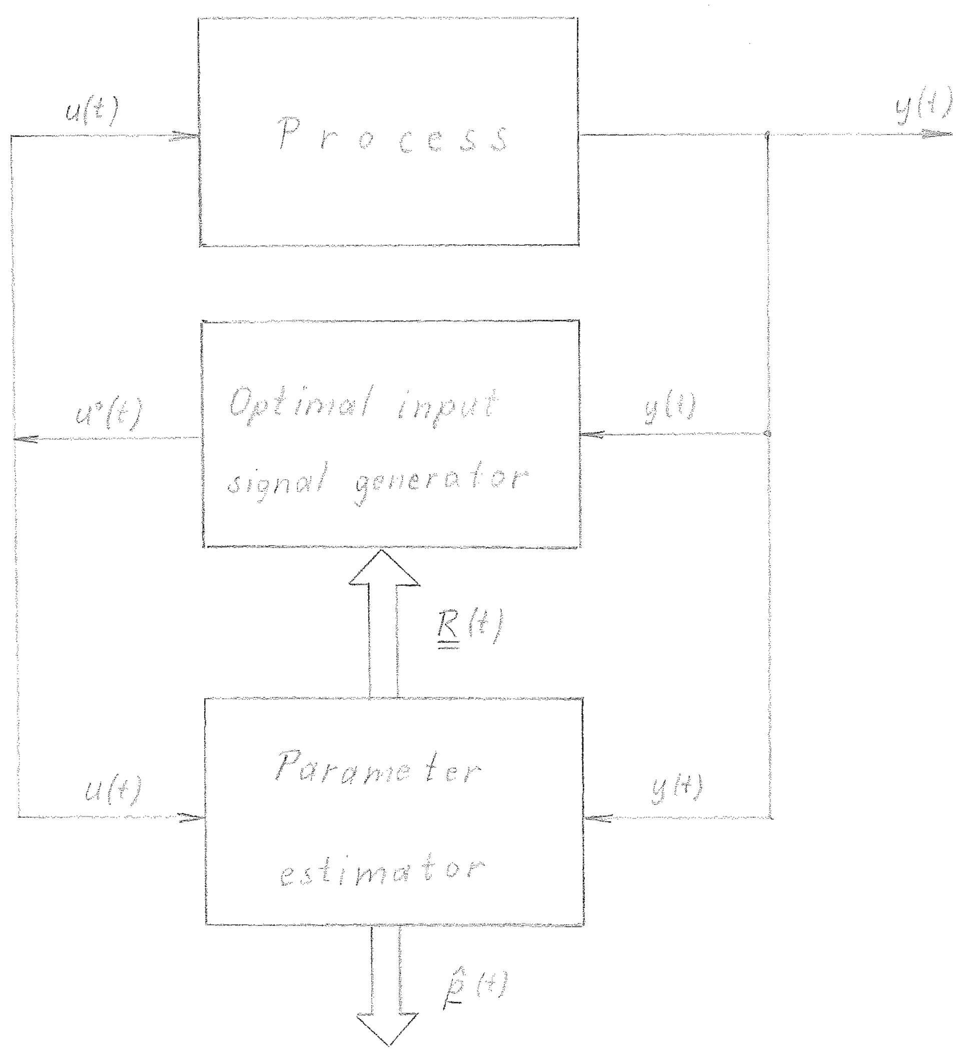

u(t)
Process
y(t)
u*(t)
Optimal input
signal generator
y(t)
R(t)
u(t)
Parameter
estimator
y(t)
p̂(t)

Fig. 5.

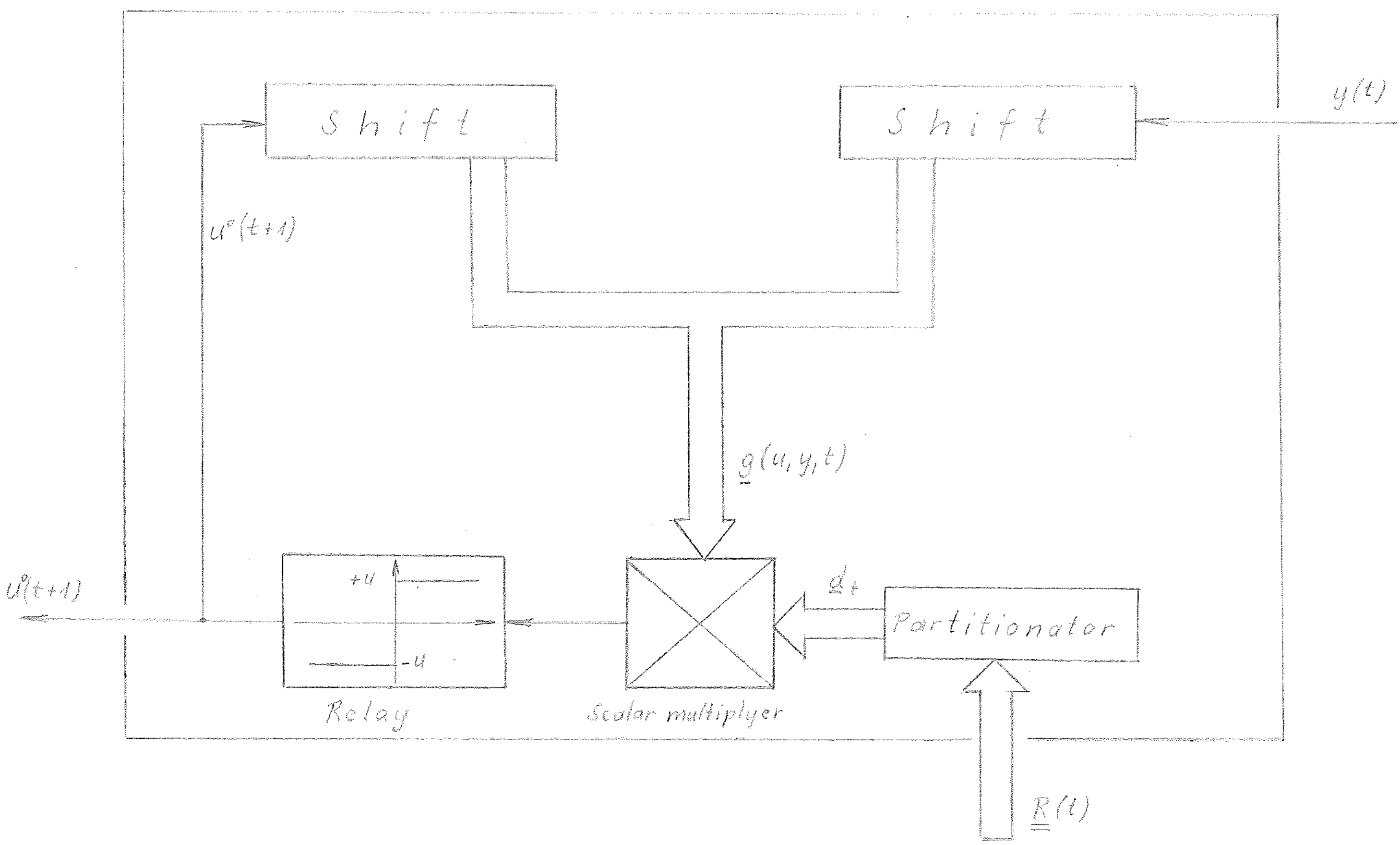

Shift
Shift
y(t)
u°(t+1)
g(u,y,t)
+u
-u
d_t
Partitionator
Relay
Scalar multiplyer
R(t)

Fig. 6.

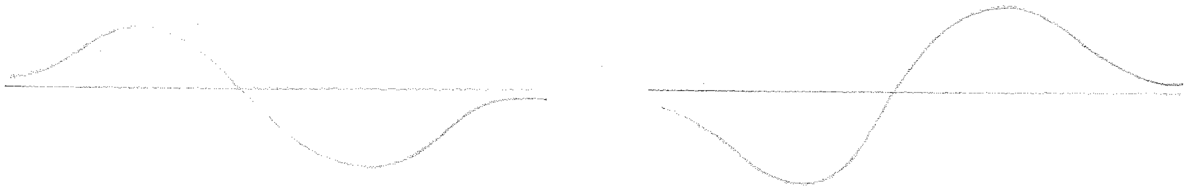


Fig. 7.

i.e. in order to produce the optimal input the global maximum of a second order rational fractional function is to be sought under the (5.1.5) restriction.

The possible forms of this function can be seen on Fig.7.(In present case the form with breaking is excluded because that is impossible physically.) According to this the maximum under such restrictions can be obtained by a numerical method. In the case without restrictions,however,the computation of $u^{o}(N+1)$ is not too complicated though it is more difficult than (5.1.14).

It is to be noted in connection with above-mentioned since now (5.9) is

$$y(t) = \underline{f}^{T}(u,y,t)\,\underline{p} + \lambda e(t) = y_{o}(t) + \lambda e(t) \qquad (5.1.17)$$

it can be seen that the synthesis procedures treated above are supported by the sensitivity function

$$\underline{s}' = \frac{dy_{o}}{d\underline{p}} = \underline{f}(u,y,t) \qquad (5.1.18)$$

so they make the system sensitive to a fictive output.

Since $E\{y\} = x$,so

$$\underline{s} = \frac{dx}{d\underline{p}} = \underline{f}(u,x,t) = E\{\underline{s}'\} = E\{\,\underline{f}(u,y,t)\} \qquad (5.1.19)$$

where

$$\underline{f}(u,x,t) = [u(t),\ldots,u(t-m),-x(t-1),\ldots,-x(t-n)]^{T} \qquad (5.1.20)$$

It can be seen from the results of Appendix C that in the case of LS structure when $\underline{\underline{C}} = \underline{\underline{I}}$ the information matrix is:

$$\underline{\underline{M}}_{N} = \frac{1}{\lambda^{2}} \sum_{t=1}^{N} \underline{f}(u,x,t)\,\underline{f}^{T}(u,x,t) \qquad (5.1.21)$$

It follows that the input synthesis procedures treated above can be applied in the case of other criteria formulated to the information matrix only $\underline{f}(u,y,t)$ must be substituted for $\underline{f}(u,x,t)$ in every equation and since the output without noise is not known it is to be produced by prediction as a model output:

$$\hat{x}(t) = \sum_{i=0}^{m} \hat{b}_i \; u(t-i) - \sum_{i=1}^{n} \hat{a}_i \; x(t-i) \qquad (5.1.22)$$

where $\hat{a}_i$ and $\hat{b}_i$ are the estimated parameters. So Fig.5. and 6. are changed according to the Figs. 8. and 9.

Since in this case (5.1.22) also has to be applied to the input signal synthesis therefore in the case of LS structure both the recursive estimation of the parameters and the real-time simulation of the process are needed.

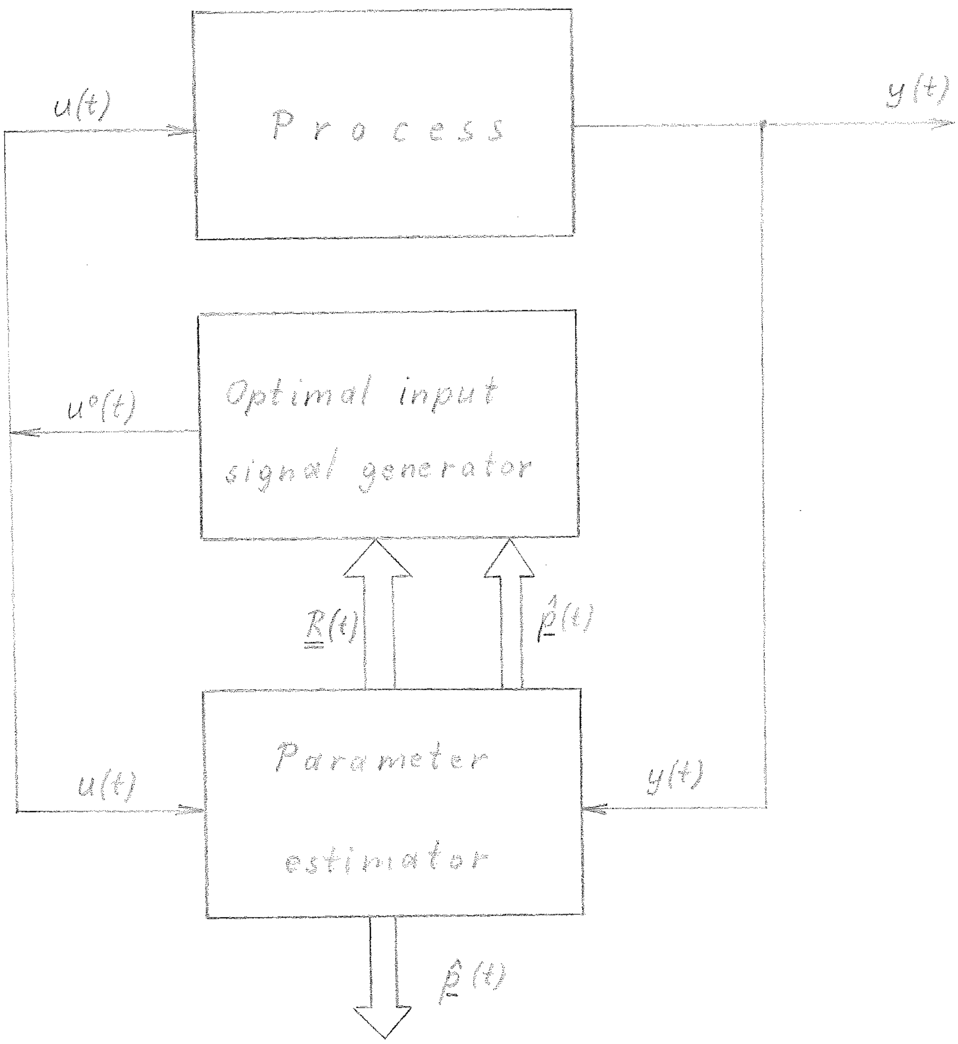

Process
u(t)
y(t)
Optimal input
signal generator
u°(t)
R(t)
p̂(t)
Parameter
estimator
u(t)
y(t)
p̂(t)

Fig. 8.

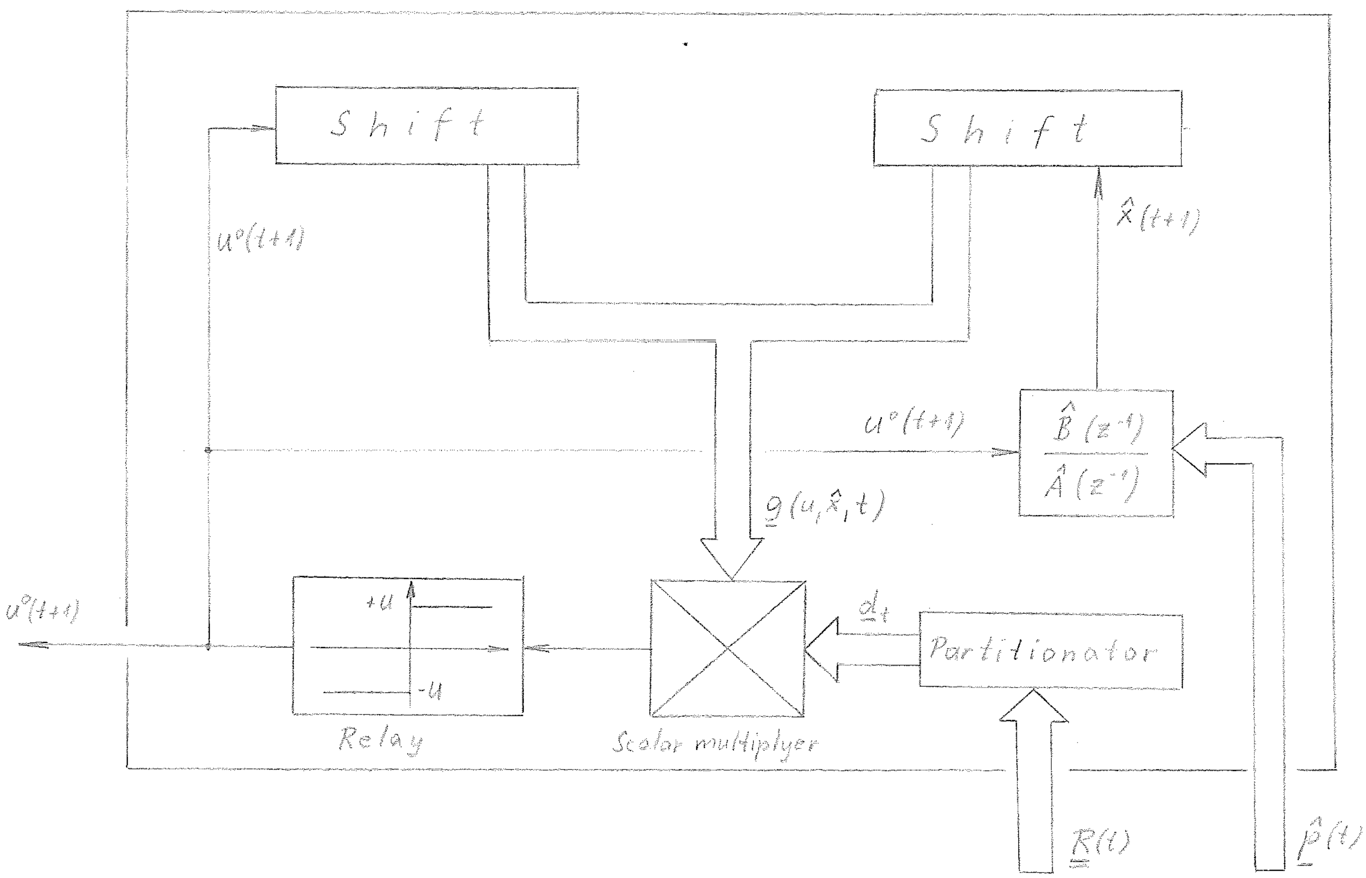

Shift
Shift
$u^o(t+1)$
$\hat{x}(t+1)$
$u^o(t+1)$
$\frac{\hat{B}(z^{-1})}{\hat{A}(z^{-1})}$
$\underline{g}(u,\hat{x},t)$
$u^o(t+1)$
+u
-u
$\underline{d}_t$
Partitionator
Relay
Scalar multiplyer
$\underline{R}(t)$
$\underline{\hat{p}}(t)$

Fig. 9.

5.2. CASE OF ÅSTRÖM-BOHLIN STRUCTURE

If our system equation corresponds to (5.1) then the information matrix according to the equation (C.9) in Appendix C is:

$$\underline{\underline{M}} = \frac{1}{\lambda^2} \sum_{t=1}^{N} \underline{f}(u^F, x^F, t)\ \underline{f}^T(u^F, x^F, t) \qquad (5.2.1)$$

where

$$u^F(t) = \frac{1}{C(z^{-1})} u(t) = u(t) - \sum_{i=1}^{n} c_i\ u^F(t-i) \qquad (5.2.2)$$

and

$$x^F(t) = \frac{1}{C(z^{-1})} x(t) = x(t) - \sum_{i=1}^{n} c_i\ x^F(t-i) \qquad (5.2.3)$$

and

$$x(t) = \sum_{i=0}^{m} b_i\ u(t-i) - \sum_{i=1}^{n} a_i\ x(t-i) \qquad (5.2.4)$$

Since (5.2.1) corresponds to (5.1.2) formally the input synthesis method deduced for the LS structure can also be applied in this case substituting $\underline{f}(u,y,t)$ for $\underline{f}(u^F,x^F,t)$ and (5.2.4) is used to produce the output without noise of the process by the estimated parameters, i.e.

$$\hat{x}(t) = \sum_{i=0}^{m} \hat{b}_i\ u(t-i) - \sum_{i=1}^{n} \hat{a}_i\ \hat{x}(t-i) \qquad (5.2.5)$$

Furthermore because of the filtering (5.2.2) now

$$\underline{f}^T(u^F, x^F, N+1) = \left[u(N+1) - q,\ \underline{g}^T(u^F, x^F, N)\right] \qquad (5.2.6)$$

where

$$q = \sum_{i=1}^{n} c_i\ u^F(t-i) \qquad (5.2.7)$$

So the rule of optimal input synthesis instead of (5.1.14) is:

$$u^{o}(N+1) = \begin{cases} +U & \text{if} \quad \underline{g}^{T}(N)\,\underline{d}_{N} > q\,r_{N} \\ -U & \text{if} \quad \underline{g}^{T}(N)\,\underline{d}_{N} \leq q\,r_{N} \end{cases} \tag{5.2.8}$$

since now the minimum point of the parabola is:

$$u^{*}(N+1) = -\frac{\underline{g}^{T}(N)\,\underline{d}_{N}}{r_{N}} + q \tag{5.2.9}$$

Note that now r_N and $\underline{d}_N$ is obtained by the partitionation of $\underline{\underline{G}} = \lambda^2 \underline{\underline{M}}_N$ and the updating of $\underline{\underline{G}}_N^{-1}$ by $\underline{f}(u^F, x^F, N+1)$ is done according to the (5.1.7).

Only it is the problem that there is no a simple and effective recursive method to estimate $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$. Therefore in the case of correlated noise off-line input synthesis method is suggested because there are good off-line identification methods (for example Åström-Bohlin method [4]).

5.3. OFF-LINE INPUT SIGNAL SYNTHESIS

If the input signal synthesis is directed to the covariance matrix then the on-line connection with the process is needed by all means since the algorithm requires the knowledge of $y(t)$. If the input signal synthesis modifies the feature of information matrix, however, the measured output is not needed only the estimates of process parameters which are used at the prediction as it was mentioned above.

So that is unconcerned how many inputs are synthetised in the possession of estimated parameters.
Thus using the parameter estimates of an off-line identification there is possibility to generate an optimum input series to the next off-line identification measurements. In this case, for example, the synthetization rule (5.2.8) can be applied for the model (5.1).

It must be emphasized that the strategies formed in this manner are only locally optimal strategies for off-line case, too, just as the previous ones mentioned in Section 5.1. and 5.2. This can also be seen from that the optimal $u^o(N+1)$ depends on even the previous values of input both in (5.1.14) and (5.2.8).
Thus if we want to generate optimal input signals for N samplings then we have to take into account the all values $u(t)$ $(t=1,\dots,N)$ simultaneously. These problems completely corresponds to the problems of "one stage" and "N-stage" strategies of discrete-time control [2,13].

For the determination of "globally" optimal input $\underline{u}$ a method was suggested by AOKI and STALEY which maximizes the trace of the information matrix, since in the case of $N \to \infty$ this strategy minimizes the trace of inverse of the information matrix for systems with limited input, output signals [12].

Applying this method for the system (5.1) on the basis of equations (C.7) and (C.8) in Appendix C it can be written

$$\mathrm{tr}(\underline{\underline{M}}_N) = \frac{1}{\lambda^2} \underline{u}^T (\underline{\underline{C}}^{-1})^T \left[\sum_{i=0}^{m} (\underline{\underline{S}}^T)^i \underline{\underline{S}}^i + (\underline{\underline{B}}\underline{\underline{A}}^{-1})^T \left(\sum_{i=1}^{n} (\underline{\underline{S}}^T)^i \underline{\underline{S}}^i \right) \underline{\underline{A}}^{-1} \underline{\underline{B}} \right] \underline{\underline{C}}^{-1} \underline{u} =$$

$$= \underline{u}^T \underline{\underline{P}}\, \underline{u} \qquad (5.3.1)$$

Thus in the case of this criterion the "globally" optimum input series $\underline{u}^o$ can be obtained by the seeking of maximum of quadratic form (5.3.1). When the restriction for u(t) is of type (5.15) (i.e. restriction in amplitude) then the extremum seeking is difficult enough. If the restriction is of power type

$$\underline{u}^T \underline{u} \leq N \lambda_u^2 \qquad (5.3.2)$$

then the solution easier since in this case the optimal input $\underline{u}^o$ is the eigenvector corresponding to the largest eigenvalue of $\underline{\underline{P}}$. Simple numerical methods are known for the computation of this latter.

6. SIMULATION RESULTS

6.1. INVESTIGATION OF ON-LINE INPUT SIGNAL SYNTHESIS

First simulation results are shown for on-line input signal synthesis by real-time simulation. These results were obtained by program INPUT made for computer UNIVAC 1108. The program performs the simulation and recursive identification of a discrete-time system of LS structure while there is possible to generate optimum input signal. Identification results obtained by PRBS and D-optimum input series (which maximizes the determinant of the information matrix) were compared by the program. First an off-line identification is performed with 2n PRBS input signals in both cases then the input signal synthesis and recursive identification are continued on the basis if initial estimates obtained for the parameters and the information matrix. The same noise series e(t) was used in the case of both input signal series.

Example 6.1.1.

The system equation

$$(1 - 0.5z^{-1})\, y(t) = 0.5\,u(t) + \lambda e(t) \qquad (6.1.1)$$

The simulation results are presented in Table I. for N=500 and λ=1.0.

The values of $tr(\lambda^2 \underline{\underline{K}}^{-1}(t))$, $tr(\underline{\underline{K}}(t)/\lambda^2)$, $|\lambda^2 \underline{\underline{K}}^{-1}(t)|$ and $\gamma(t)$ are shown on Figs.10.,11.,12. and 13.,respectively as a function of time t.

Here $\gamma(t)$ characterizes the consistence of estimation:

$$\gamma(t) = \sum_{i=1}^{k} \frac{|\hat{p}_i(t) - p_i|}{|\hat{p}_i(0) - p_i|} \tag{6.1.2}$$

The determinant and trace of matrix corresponding to the CRAMER-RAO lower bound can be calculated for N=500 and $\lambda = 1.0$ on the basis of Appendix D:

$$|\underline{\underline{M}}_{500}^{-1}| = 2.4 \; 10^{-6} \tag{6.1.3}$$

and

$$tr(\underline{\underline{M}}_{500}^{-1}) = 3.2 \; 10^{-3} \tag{6.1.4}$$

For PRBS input we get from Figs.11. and 12. that

$$\frac{|\underline{\underline{K}}_{500}|}{|\underline{\underline{K}}_o|} = 5.2622 \; 10^{-4} \geq |\underline{\underline{M}}_{500}^{-1}| \tag{6.1.5}$$

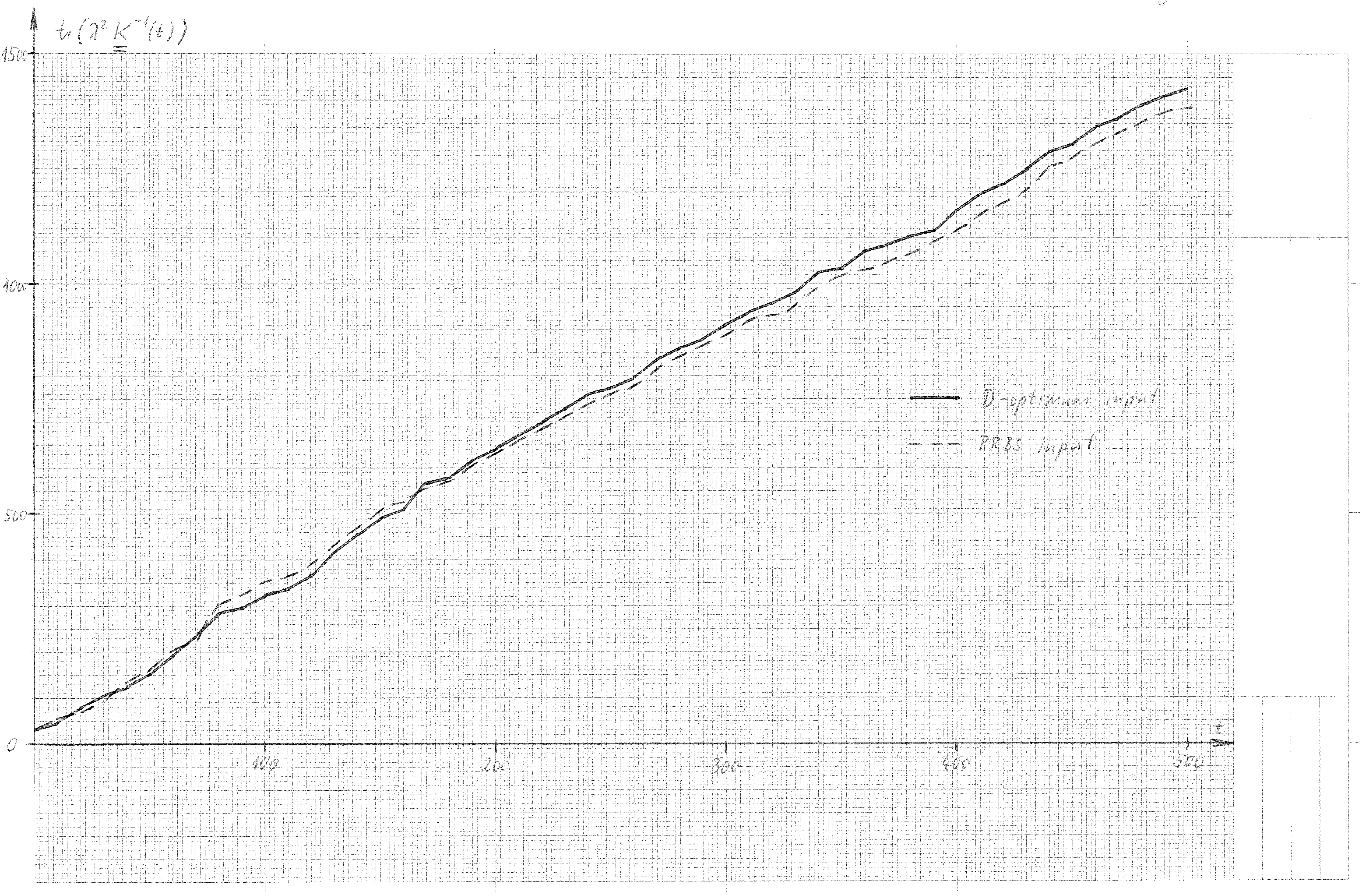
$$\frac{tr(\underline{\underline{K}}_{500})}{tr(\underline{\underline{K}}_o)} = 2.1864 \; 10^{-2} \geq tr(\underline{\underline{M}}_{500}^{-1}) \tag{6.1.6}$$

and for D-optimum input

$$\frac{|\underline{\underline{K}}_{500}|}{|\underline{\underline{K}}_o|} = 5.0837 \; 10^{-4} \geq |\underline{\underline{M}}_{500}^{-1}| \tag{6.1.7}$$

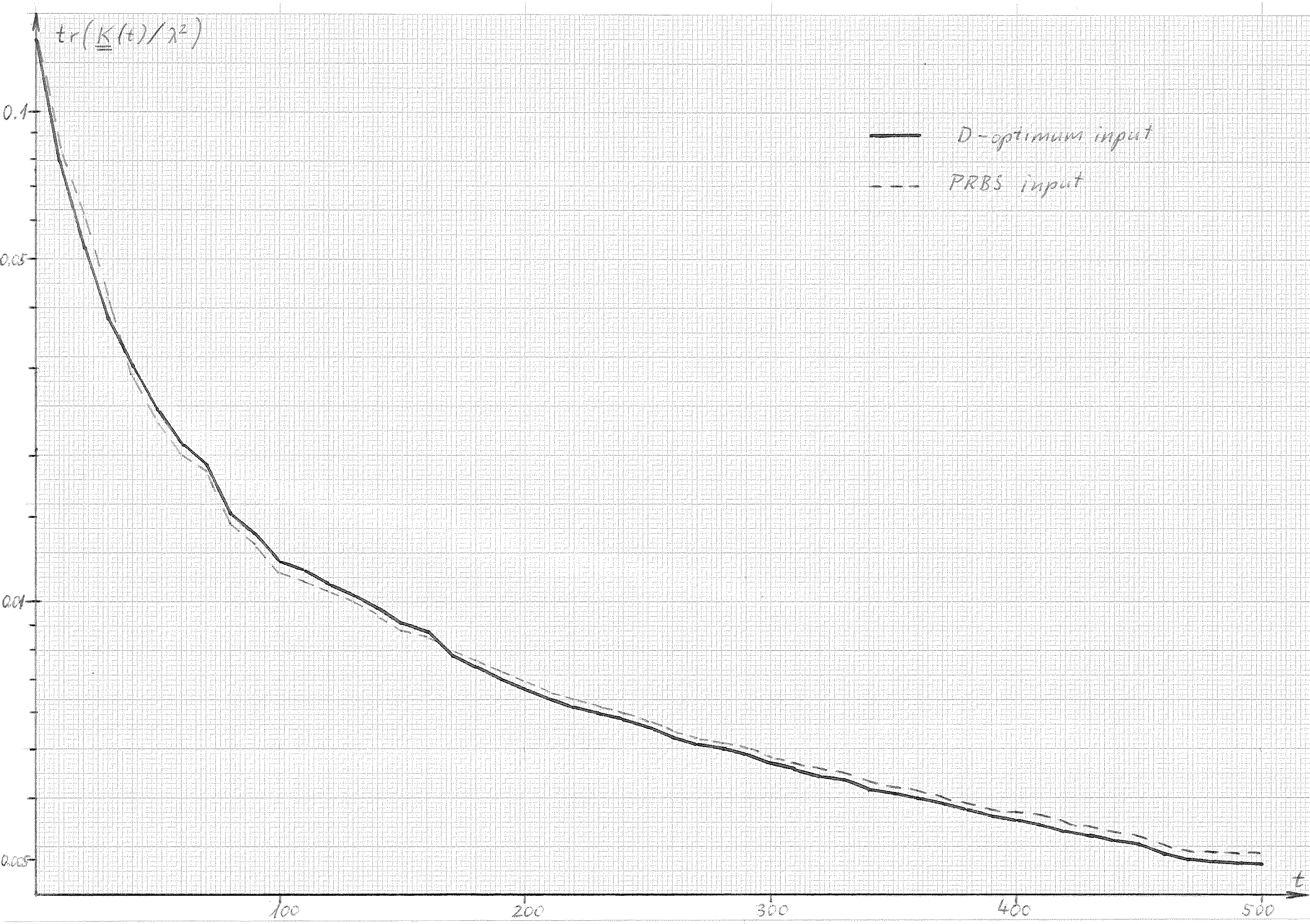
$$\frac{tr(\underline{\underline{K}}_{500})}{tr(\underline{\underline{K}}_o)} = 2.1591 \; 10^{-2} \geq tr(\underline{\underline{M}}_{500}^{-1}) \tag{6.1.8}$$

Figure 10.



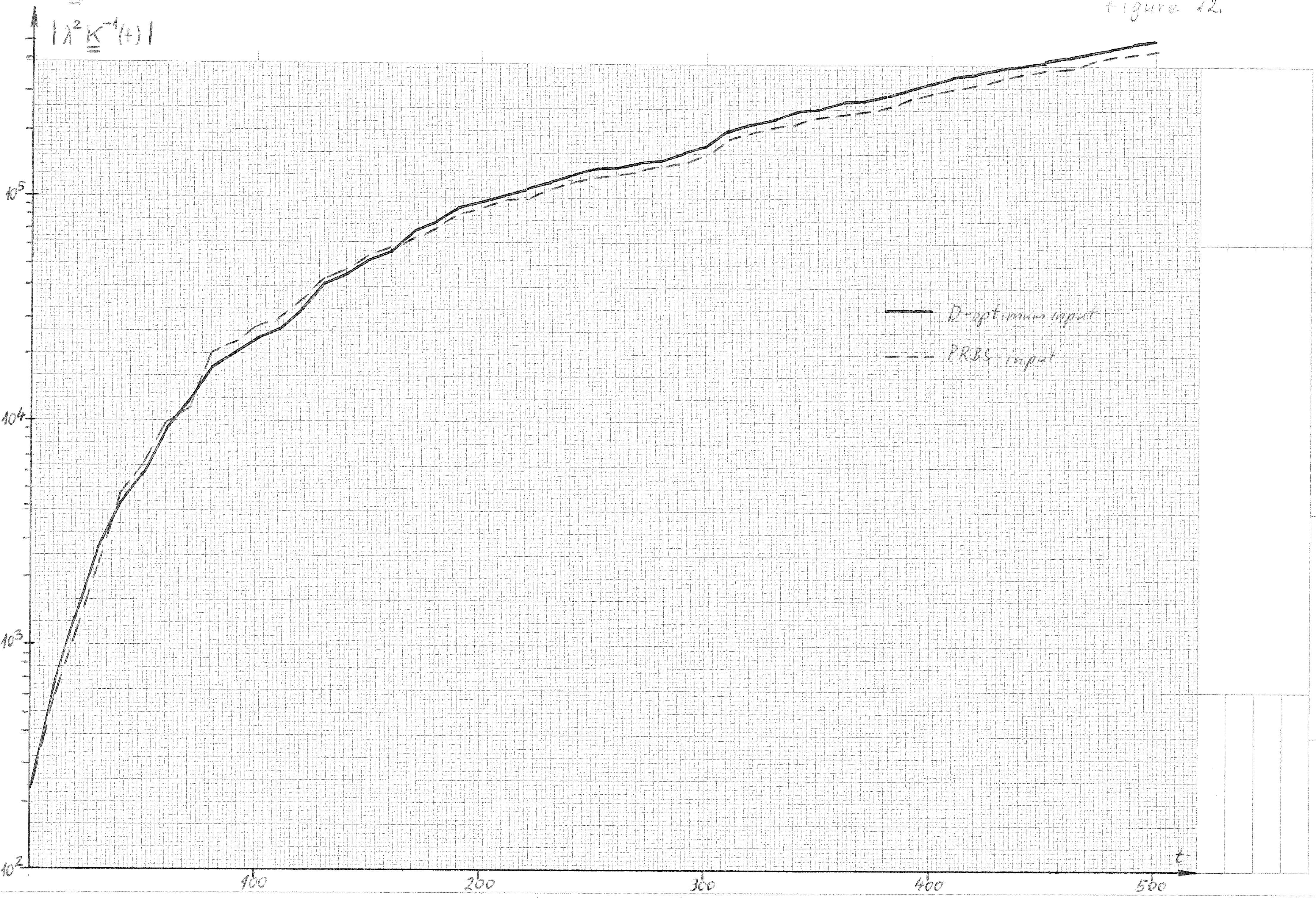
$tr(\lambda^2 \underline{K}^{-1}(t))$
1500
1000
500
0
100
200
300
400
500
t
D-optimum input
PRBS input

Figure 11.



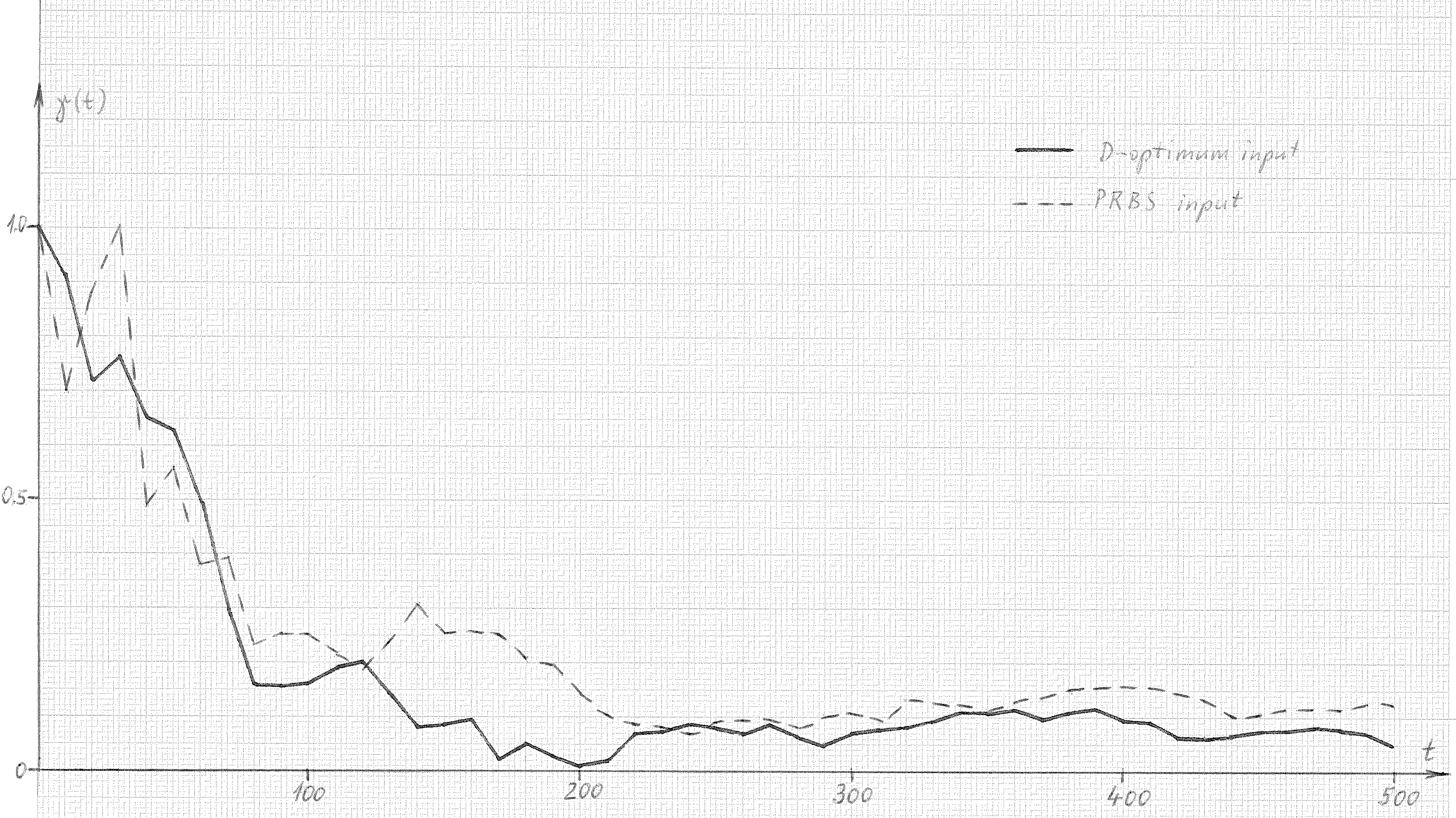
tr(K(t)/λ²)
0.1
0.05
0.01
0.005
D-optimum input
PRBS input
100
200
300
400
500
t

Figure 12.



$|\lambda^2 \underline{\underline{K}}^{-1}(t)|$
10^5
10^4
10^3
10^2
100
200
300
400
500
t
D-optimum input
PRBS input

Figure 13.



It can be seen that the approaching of CRAMER-RAO lower bound is better in the case of D-optimum input than PRBS signal but the difference is not too large.

Example 6.1.2.

The system equation

$$(1 - 1.5z^{-1} + 0.7z^{-2})\, y(t) = (1 + 0.5z^{-1})\, u(t) + \lambda e(t) \qquad (6.1.9)$$

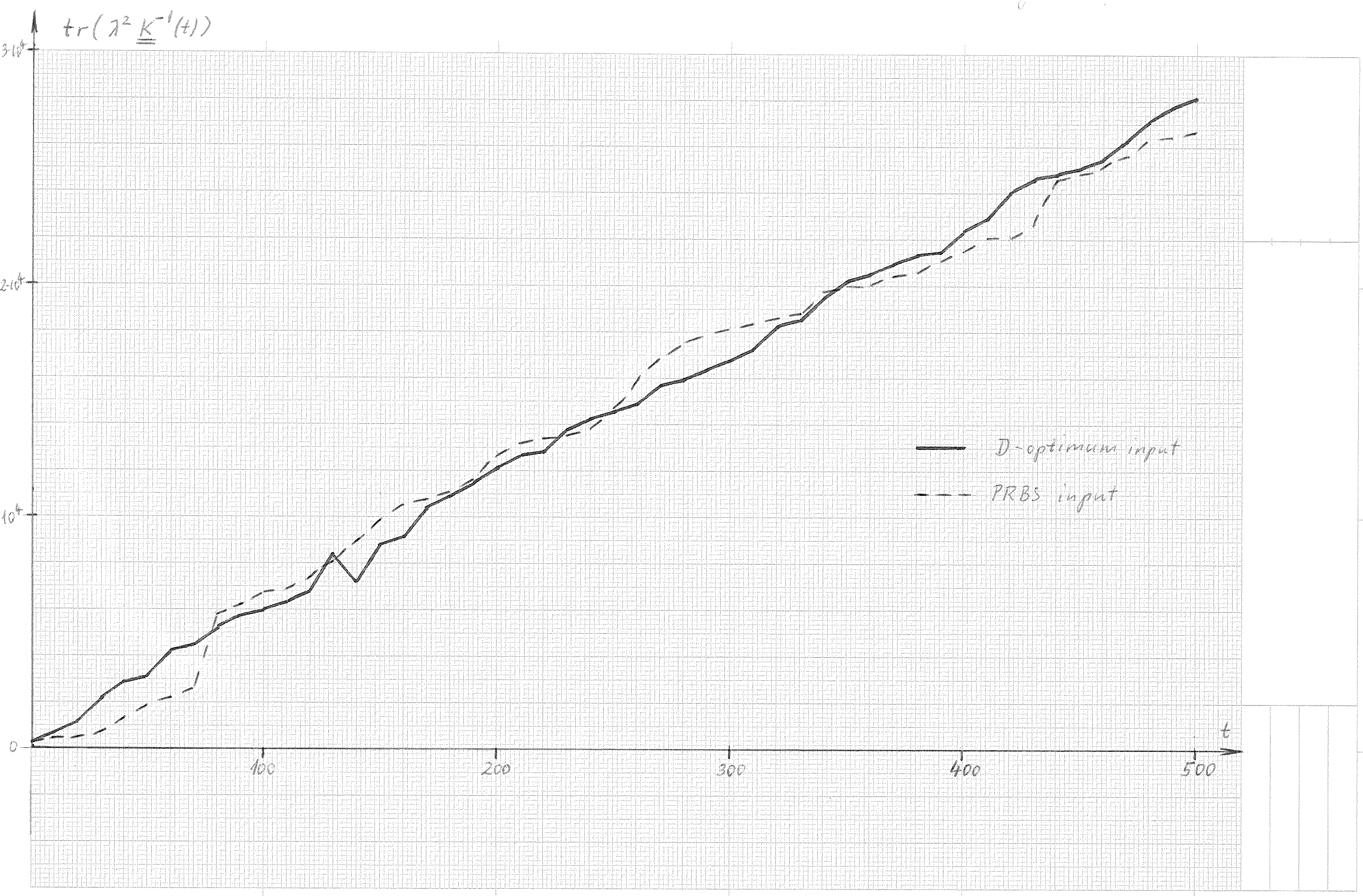
The simulation results are shown in Table II. where N=500, λ=1.0.
The values of $tr(\lambda^2 \underline{\underline{K}}^{-1}(t))$, $tr(\underline{\underline{K}}(t)/\lambda^2)$, $|\lambda^2\underline{\underline{K}}^{-1}(t)|$ and $\gamma(t)$ are presented on Figs. 14.,15.,16.,17.,respectively.

Example 6.1.3.

System equation

$$(1 - 1.7z^{-1} + 1.0z^{-2} - 0.14z^{-3})\, y(t) = (1 + 0.5z^{-1} - 0.5z^{-2} + 0.1z^{-3})\, u(t) + \lambda e(t) \qquad (6.1.10)$$

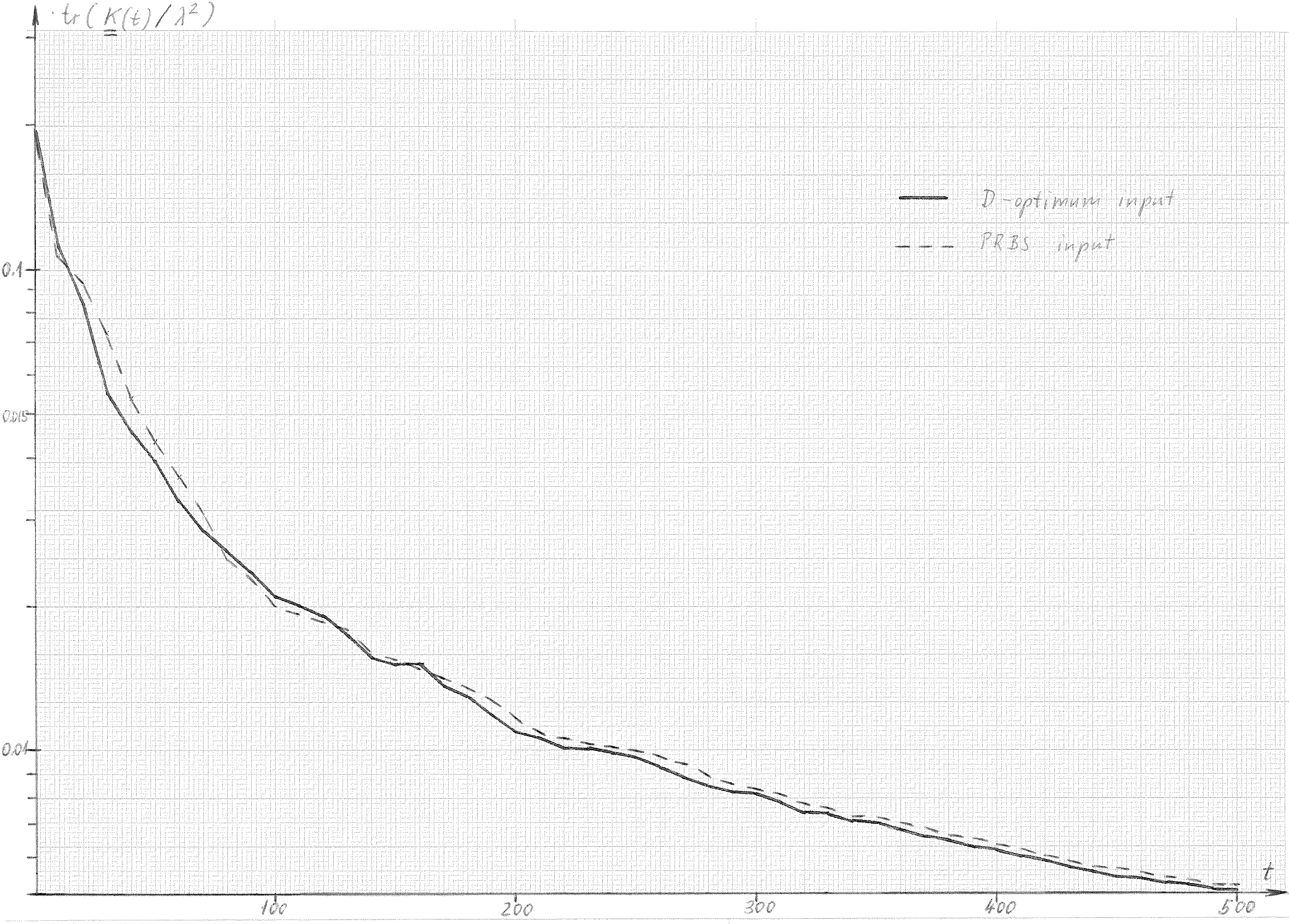
The simulation results can be seen in Table III. for N=100 , λ=0.02.
The performance characteristics are presented on Figs.18.,19.,20.,21. as above.



tr(λ² K⁻¹(t))
3·10⁴
2·10⁴
10⁴
0
100
200
300
400
500
t
D-optimum input
PRBS input

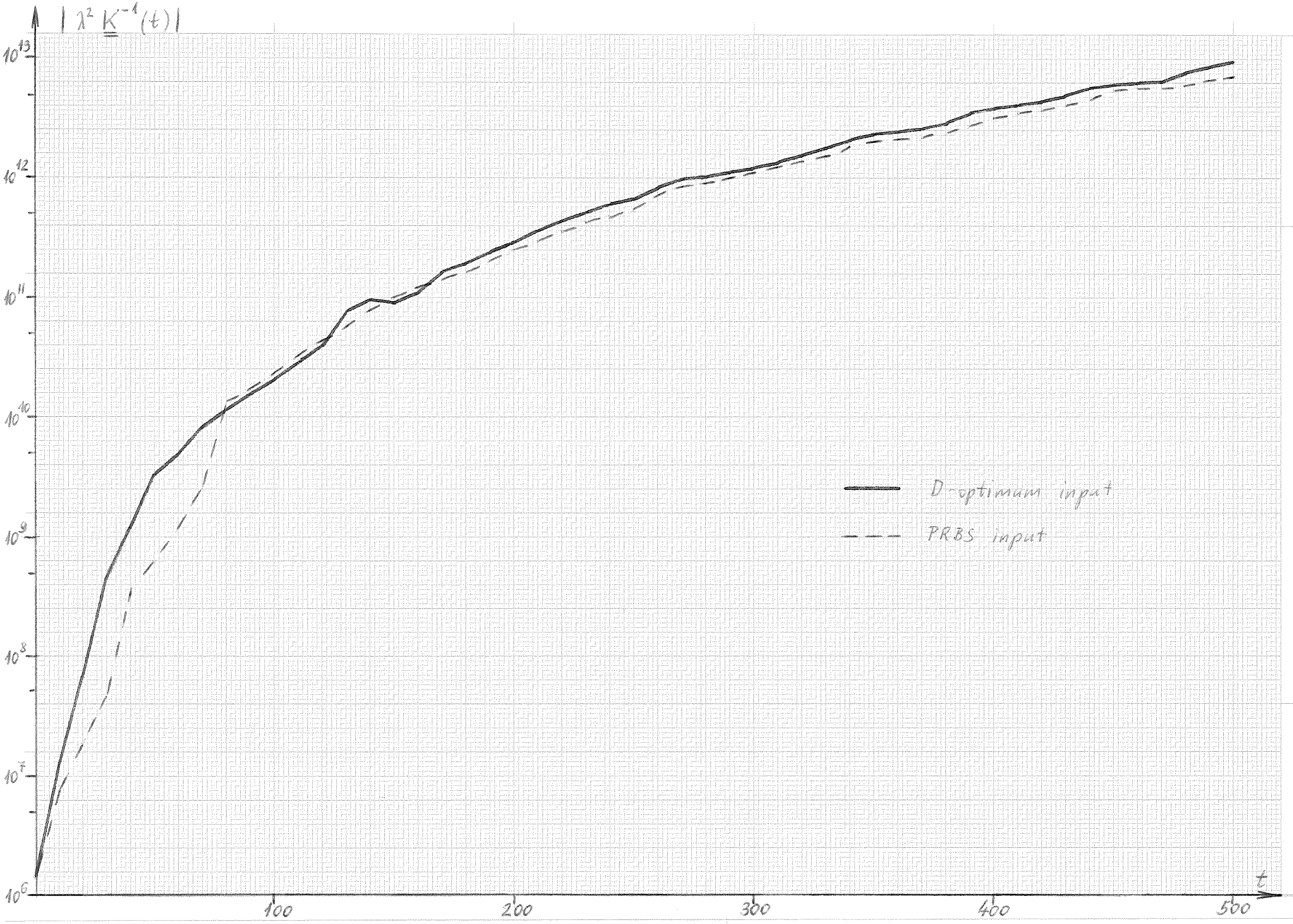
Figure 14.

Figure 15.



tr(K(t)/λ²)
0.1
0.015
0.01
D-optimum input
PRBS input
100
200
300
400
500
t

Figure 16



$|\lambda^2 \underline{K}^{-1}(t)|$
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
100
200
300
400
500
t
D-optimum input
PRBS input

Figure 17.

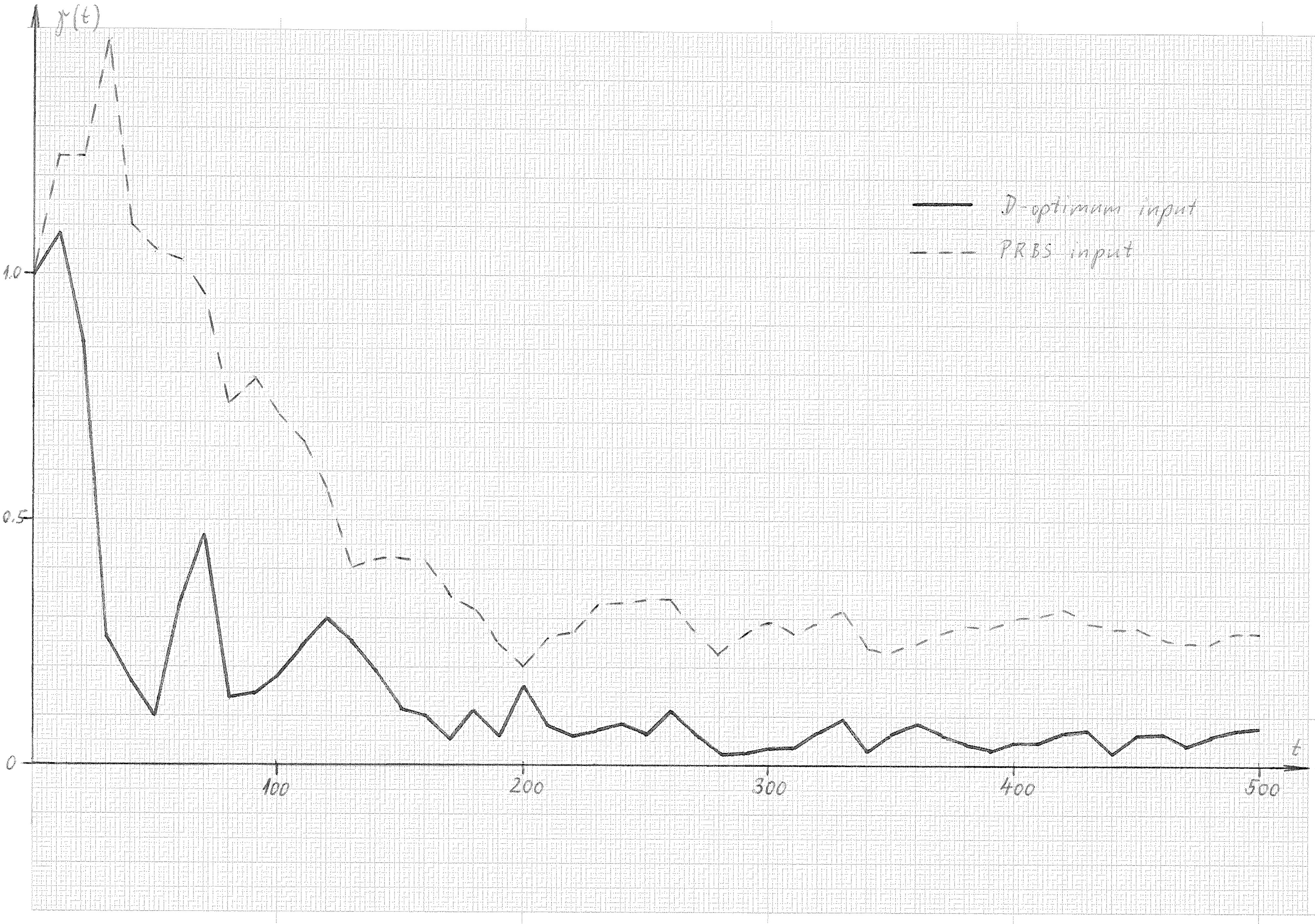
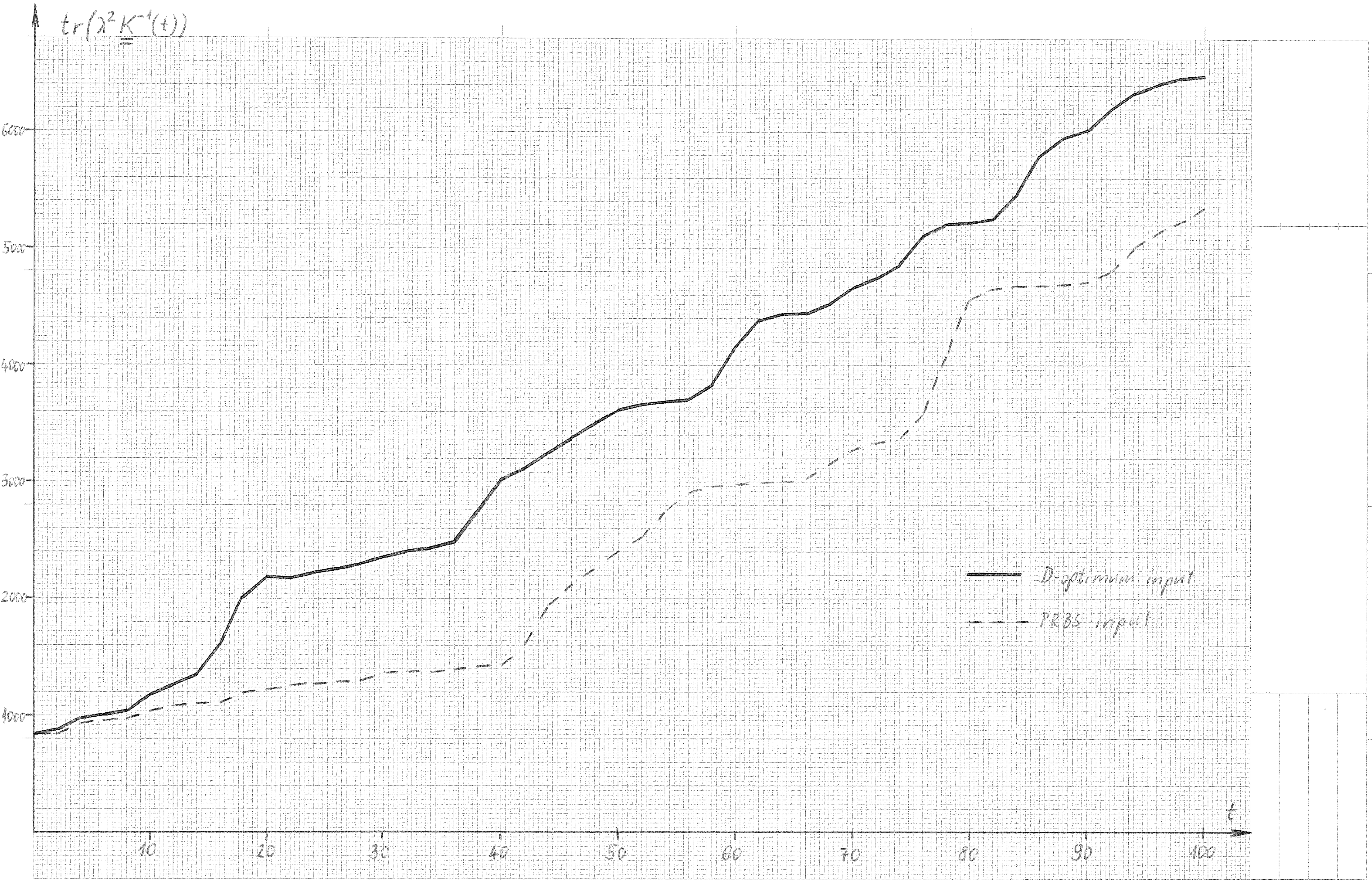
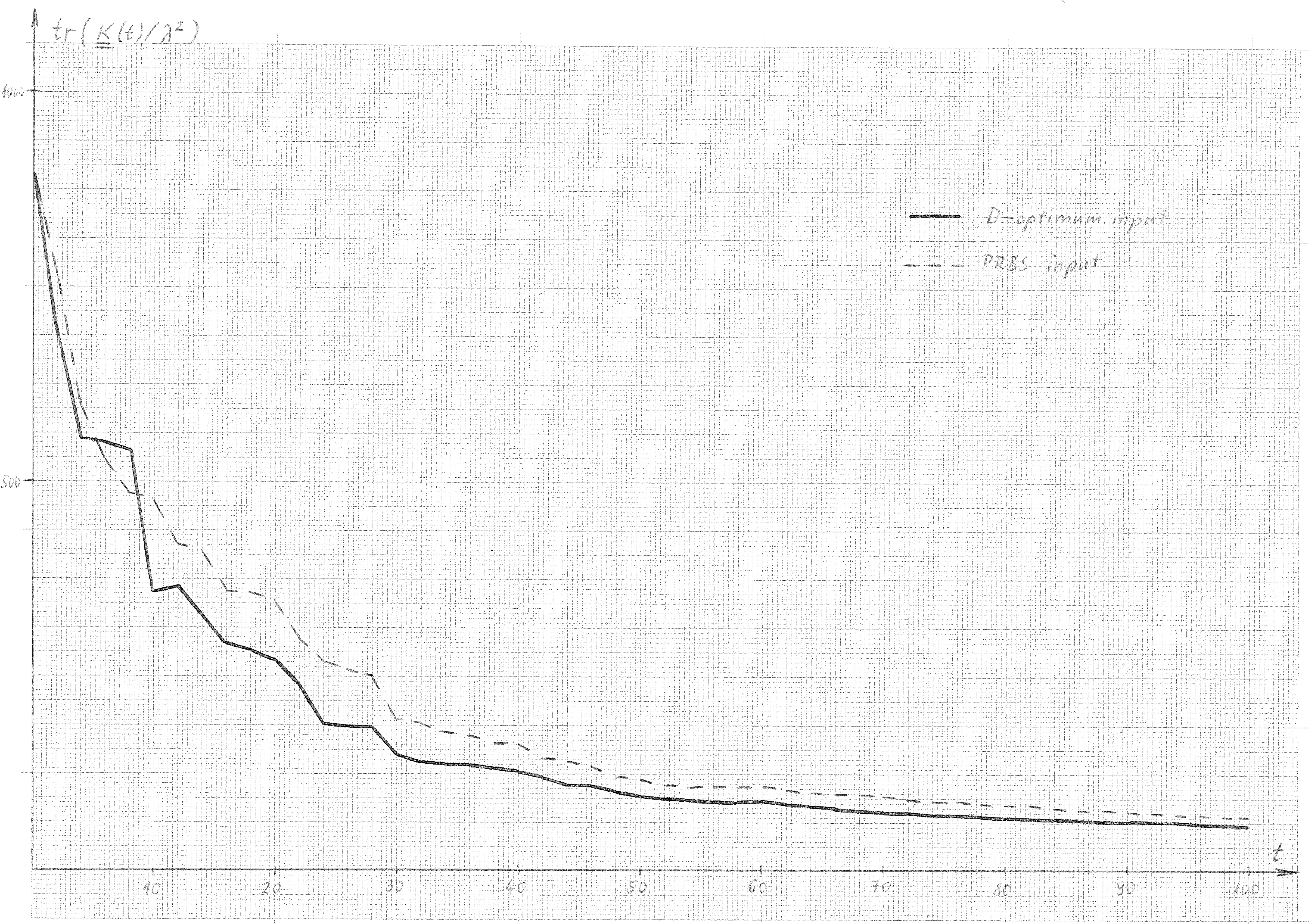


Figure 18.



$tr(\lambda^2 K^{-1}(t))$
6000
5000
4000
3000
2000
1000
10
20
30
40
50
60
70
80
90
100
t
D-optimum input
PRBS input

Figure 19



tr(K(t)/λ²)
1000
500
D-optimum input
PRBS input
t
10
20
30
40
50
60
70
80
90
100

Figure 20.

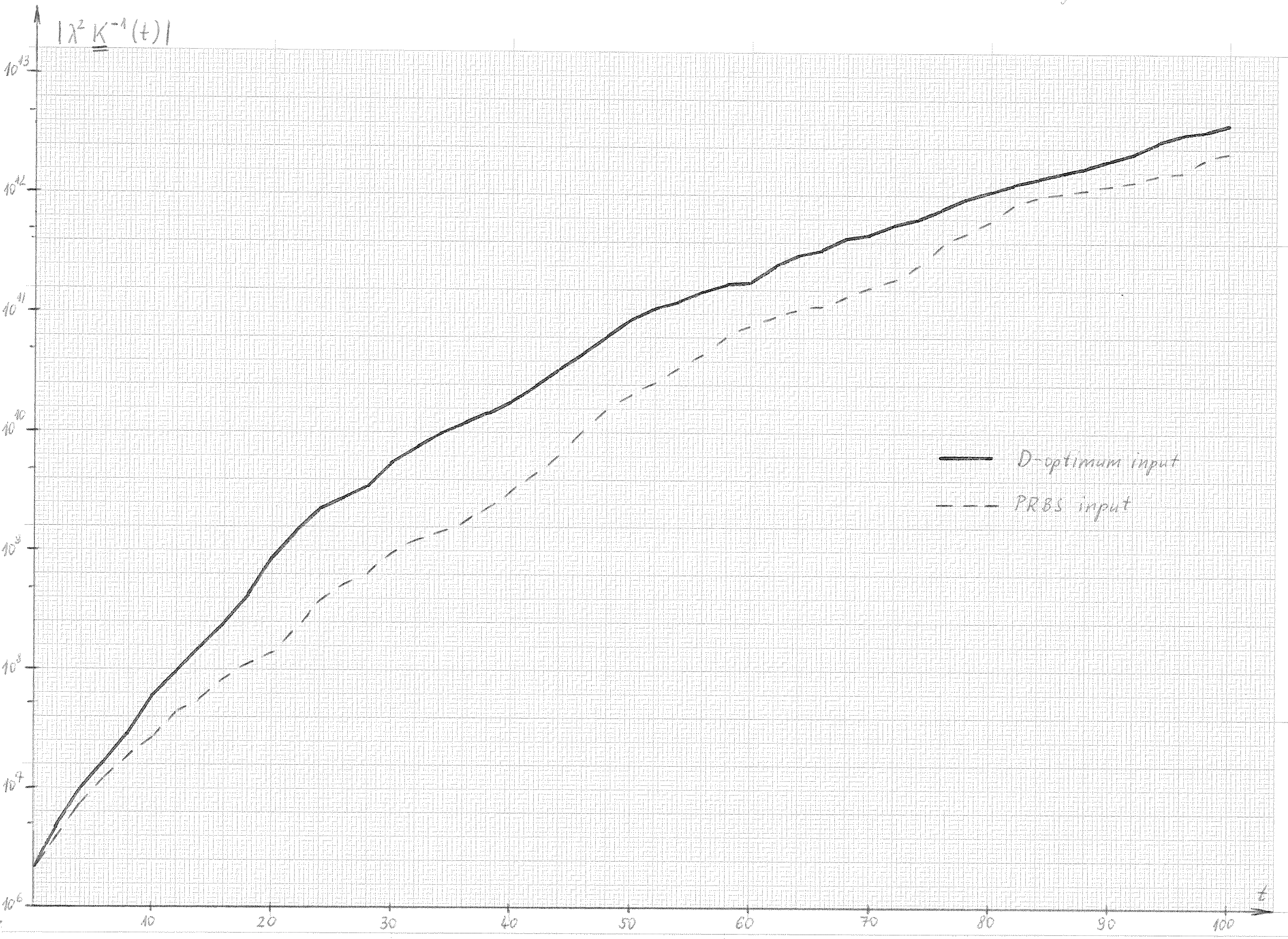
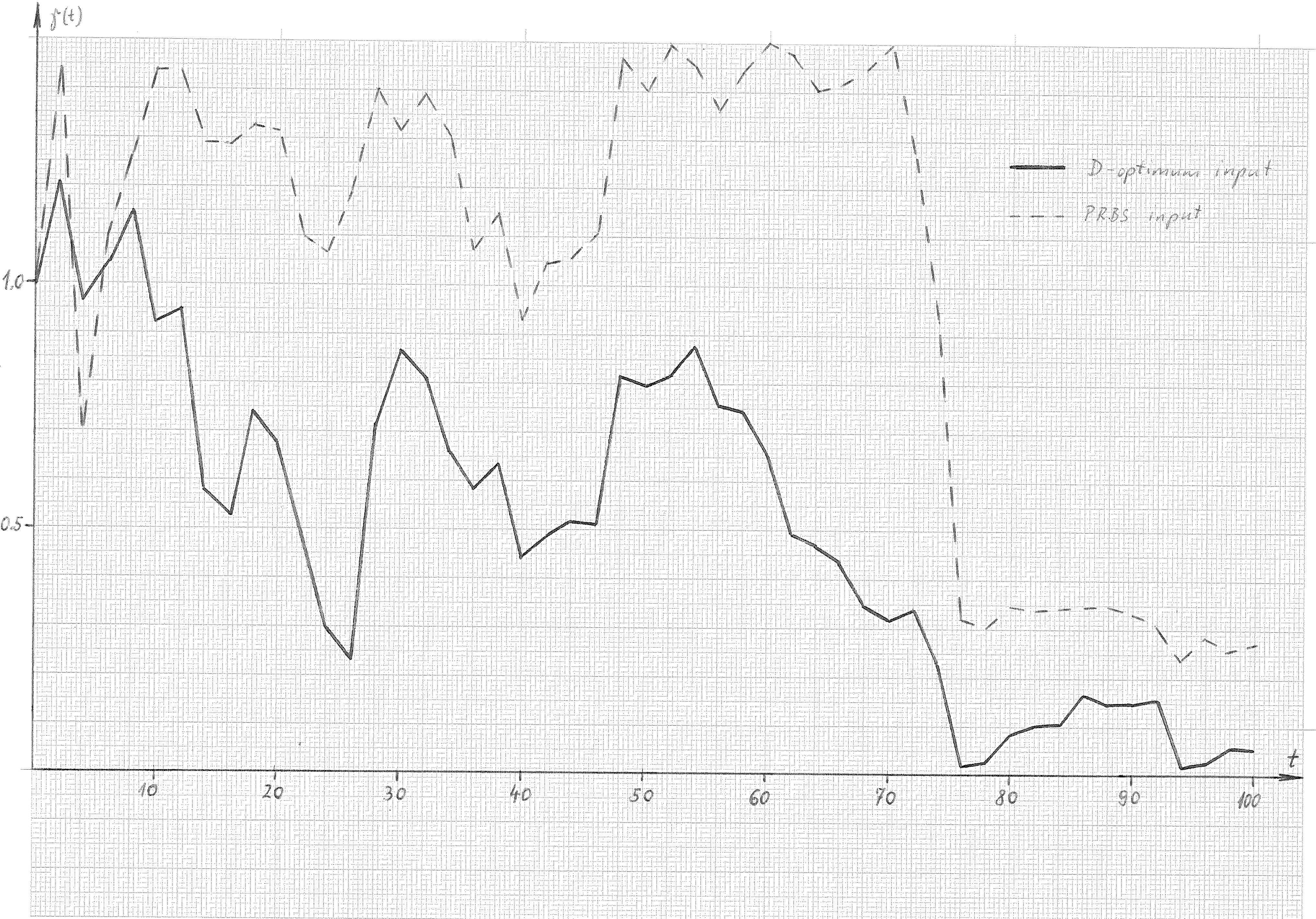

$|\lambda^2 \underline{K}^{-1}(t)|$
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10
20
30
40
50
60
70
80
90
100
t
D-optimum input
PRBS input

Figure 21.



ρ(t)
1.0
0.5
t
10
20
30
40
50
60
70
80
90
100
D-optimum input
PRBS input

Parameters	True values	Initial estimates	Final estimates	
			PRBS input	D-optimum input
a_1	-0.5	0.0355	-0.4349	-0.4526
b_0	0.5	0.3184	0.4788	0.4895

Table I. $N=500$; $\lambda=1.0$

Parameters	True values	Initial estimates	Final estimates	
			PRBS input	D-optimum input
a_1	-1.5	-1.2625	-1.4840	-1.4827
a_2	0.7	0.5021	0.6866	0.6338
b_0	1.0	0.9068	0.9765	1.0231
b_1	0.5	0.3919	0.4291	0.4925

Table II. $N=500$; $\lambda=1.0$

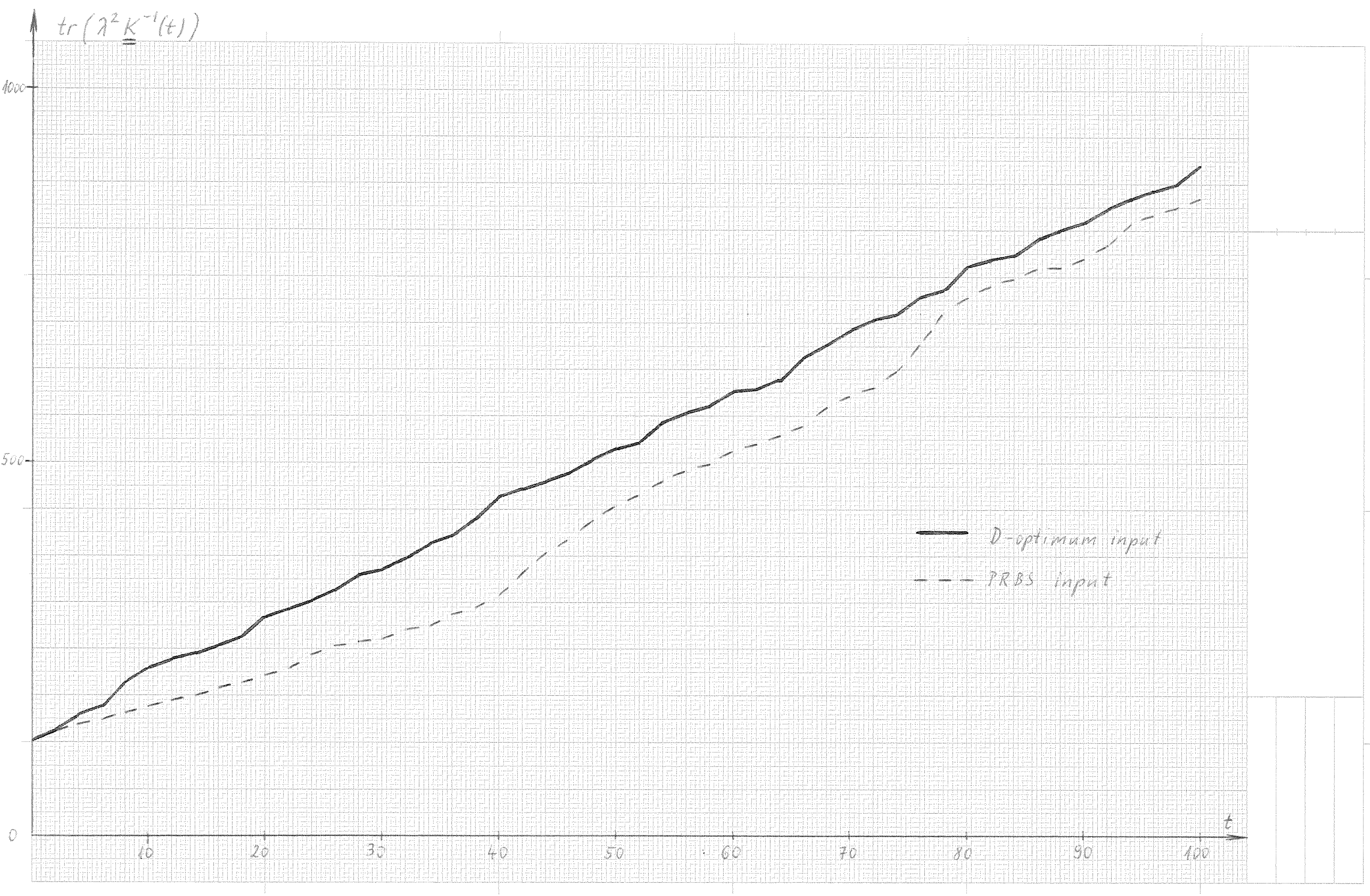
Example 6.1.4.

The system equation is:

$$(1 - 0.5z^{-1} + 0.0z^{-2})\ y(t) = (1 + 0.5z^{-1} + 0.0z^{-2})\ u(t) + \lambda e(t) \tag{6.1.11}$$

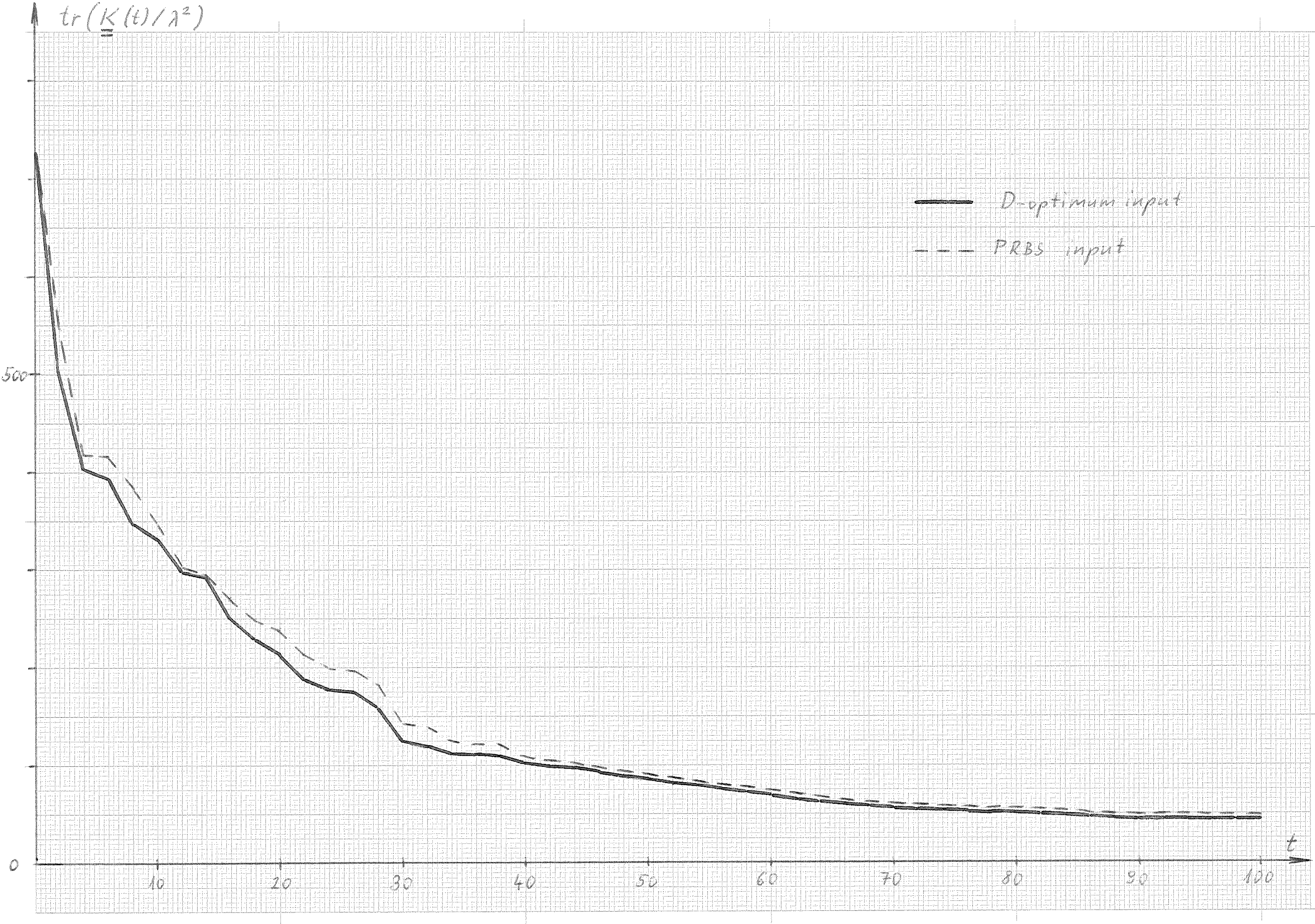
i.e. now the estimates obtained in the case of over estimated structure are presented in Table IV.
The Figs.22.,23.,24.,25. show the procedure of input signal synthesis for Eq.(6.1.11).

Figure 22.



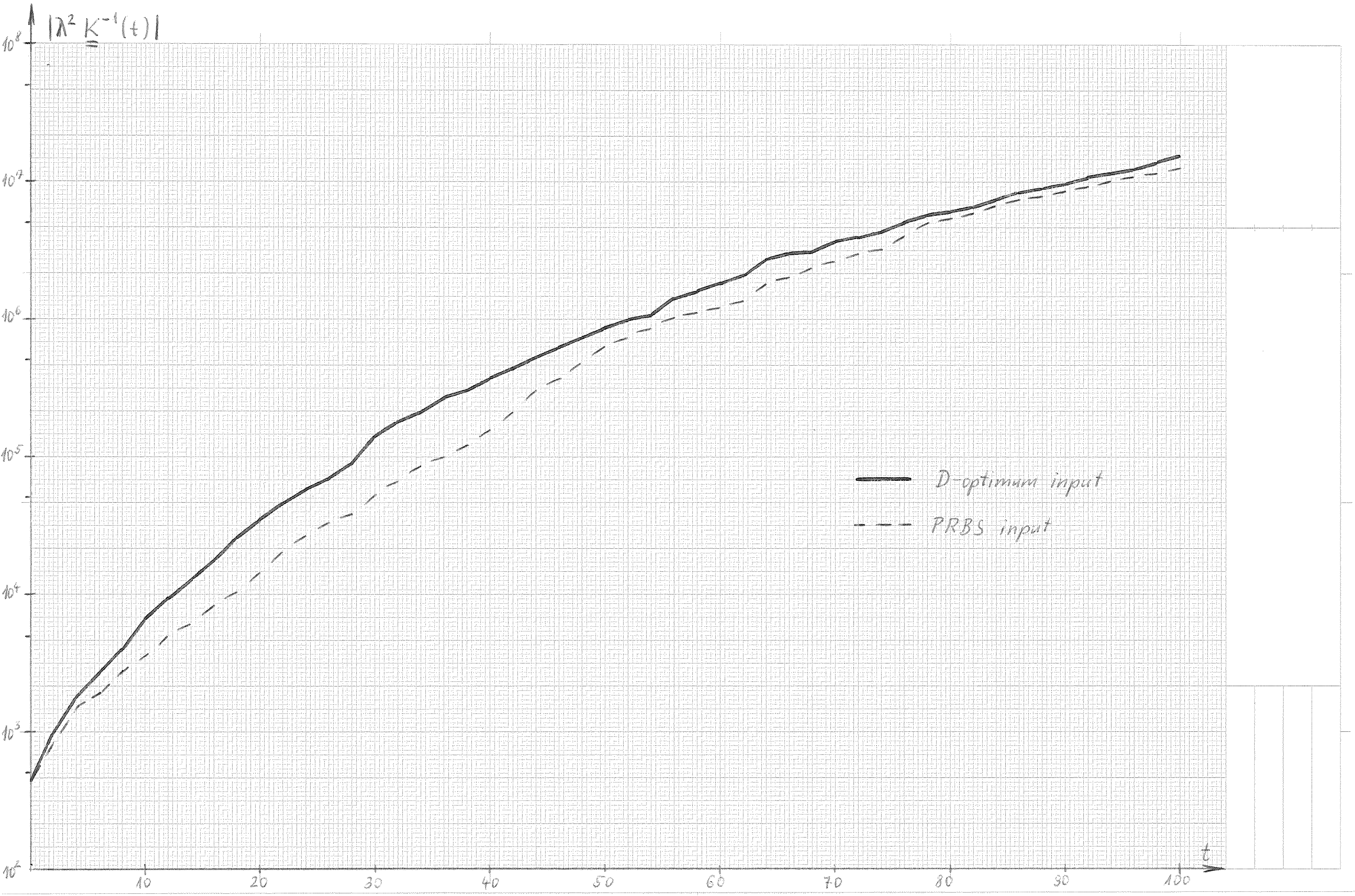
tr(λ² K⁻¹(t))
1000
500
0
t
10
20
30
40
50
60
70
80
90
100
D-optimum input
PRBS input

Figure 23.



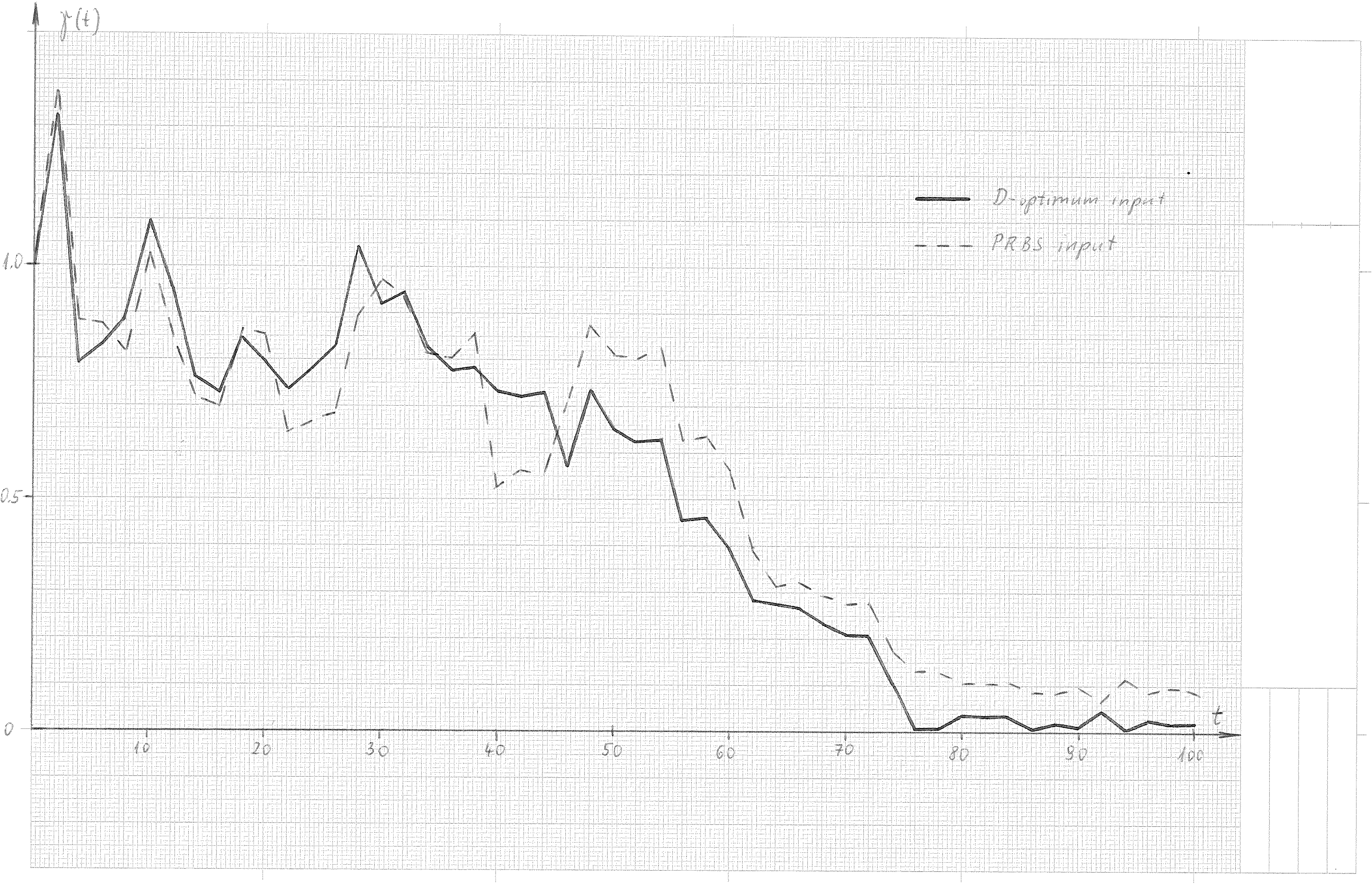
tr(K(t)/λ²)
D-optimum input
PRBS input
500
0
10
20
30
40
50
60
70
80
90
100
t

Figure 24



$|\lambda^2 \underline{K}^{-1}(t)|$
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10
20
30
40
50
60
70
80
90
100
t
D-optimum input
PRBS input

Figure 5



γ(t)
1.0
0.5
0
t
10
20
30
40
50
60
70
80
90
100
D-optimum input
PRBS input

Parameters	True values	Initial estimates	Final estimates	
			PRBS input	D-optimum input
a_1	-1.7	-1.5444	-1.6567	-1.7105
a_2	1.0	0.7648	0.9356	1.0140
a_3	-0.14	-0.0313	-0.1103	-0.1461
b_0	1.0	0.9984	0.9975	1.0007
b_1	0.5	0.6488	0.5467	0.4870
b_2	-0.5	-0.3940	-0.4701	-0.5065
b_3	0.1	0.0343	0.0842	0.1024

Table III. $N=100$; $\lambda=0.02$

Parameters	True values	Initial estimates	Final estimates	
			PRBS input	D-optimum input
a_1	-0.5	-0.1178	-0.5330	-0.4973
a_2	0.0	-0.1889	0.0179	-0.0057
b_0	1.0	0.9967	0.9976	0.9979
b_1	0.5	0.8761	0.4706	0.5070
b_2	0.0	0.1911	-0.0163	0.0008

Table IV. $N=100$; $\lambda=0.02$

6.2. INVESTIGATION OF OFF-LINE INPUT SIGNAL SYNTHESIS

We have also made some simulation investigations for off-line input signal synthesis. The system equation was (5.1), PRBS signal and D-optimum input series generated according to (5.3) and (5.2.8) were used as input signals. The number of samplings was N=500. The same series e(t) is applied in the case of both input signals. During the generating of optimum input signal the parameters of the process are assumed known. The program MISOID in the program library [7] produces the ML estimates of the parameters on the basis of simulation results.

Example 6.2.1.

The simulated process is:

$$(1 - 0.5z^{-1})\, y(t) = 0.5z^{-1}\, u(t) + \lambda\, (1 - 0.2z^{-1})\, e(t) \qquad (6.2.1)$$

The identification results can be seen in Table V. In the Table V. V means the value of loss function.

For this system on the basis of Appendix E :

$$|\underline{\underline{M}}_{500}^{-1}| = 2.2008\ 10^{-6} \quad \text{and} \quad tr(\underline{\underline{M}}_{500}^{-1}) = 3.0804\ 10^{-3}\ .$$

(Obviously here $\underline{\underline{M}}$ relates only to a_i and b_i .)

It can be seen that the CRAMER-RAO lower bound is approached much better in the case of D-optimum signal than PRBS.

Example 6.2.2.

The system equation

$$(1 - 0.5z^{-1})\, y(t) = 0.5z^{-1}\, u(t) + \lambda e(t) \qquad (6.2.2)$$

The ML estimates are shown in Table VI.

The determinant and trace of inverse of the information matrix is

$$|\underline{\underline{M}}_{500}^{-1}| = 2.4\ 10^{-6} \quad \text{and} \quad tr(\underline{\underline{M}}_{500}^{-1}) = 3.2\ 10^{-3}\ .$$

Example 6.2.3.

The simulated process:

$$(1 - 0.5z^{-1})\, y(t) = 0.5z^{-1}\, u(t) + \lambda(1 - 0.5z^{-1})\, e(t) \qquad (6.2.3)$$

The ML estimates can be seen in Table VII. , and

$$|\underline{\underline{M}}_{500}^{-1}| = 1.5577\ 10^{-6}\ , \quad tr(\underline{\underline{M}}_{500}^{-1}) = 3.0804\ 10^{-3}\ .$$

It can be established from the above that the D-optimum input series provide a ML estimation nearer to the CRAMER-RAO lower bound than the PRBS input in the case of first order system.

Now let us investigate second order systems.

Example 6.2.4.

The simulated process:

$$(1-1.5z^{-1}+0.7z^{-2})y(t) = (1.0z^{-1}+0.5z^{-2})u(t) + \lambda(1-1.8z^{-1}+0.9z^{-2})e(t) \qquad (6.2.4)$$

The ML estimates are in Table VIII. The values of input,output,model output,model error and residuals are presented on Figs.26. and 27. for PRBS and D-optimum input, respectively.

Figure 26

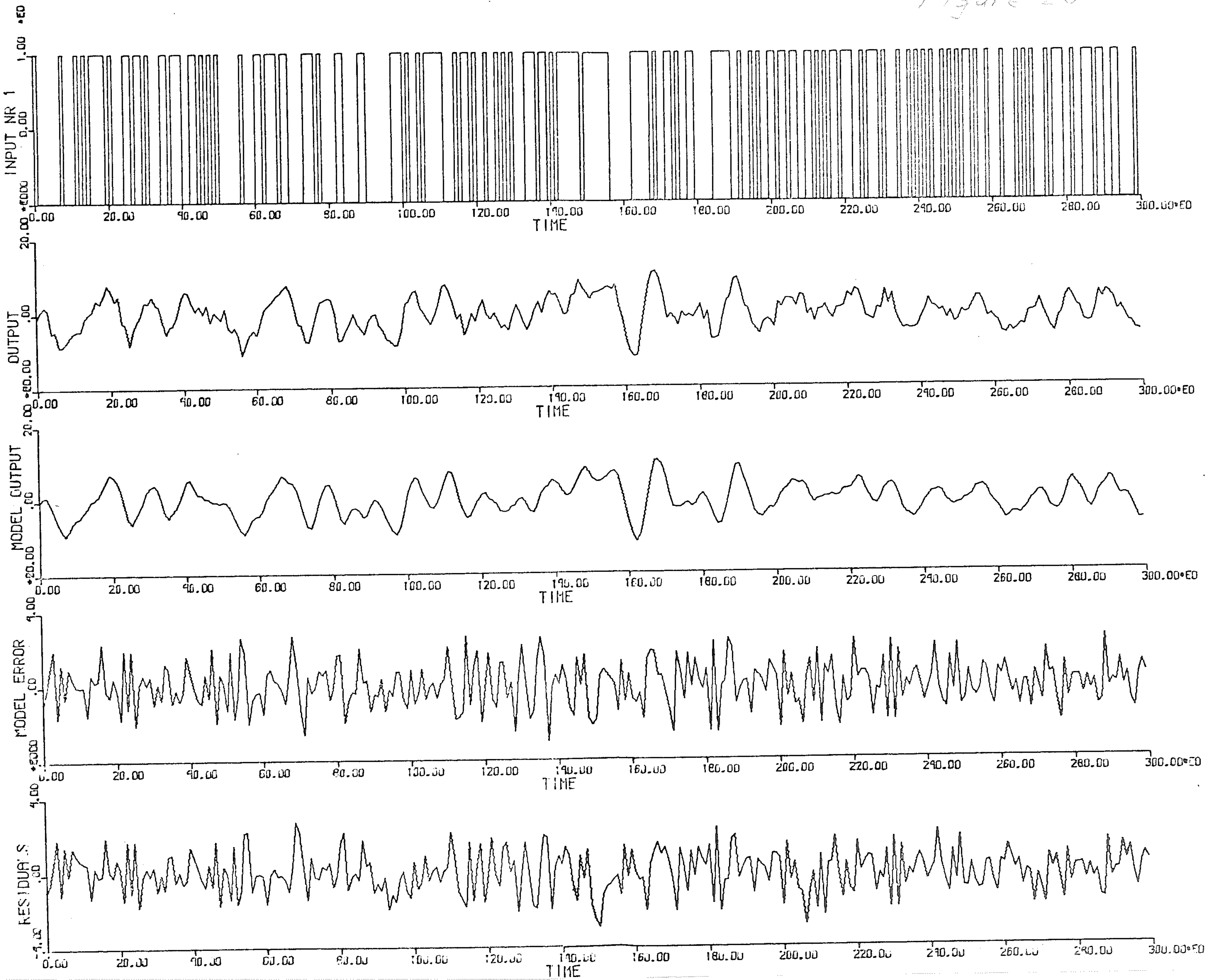
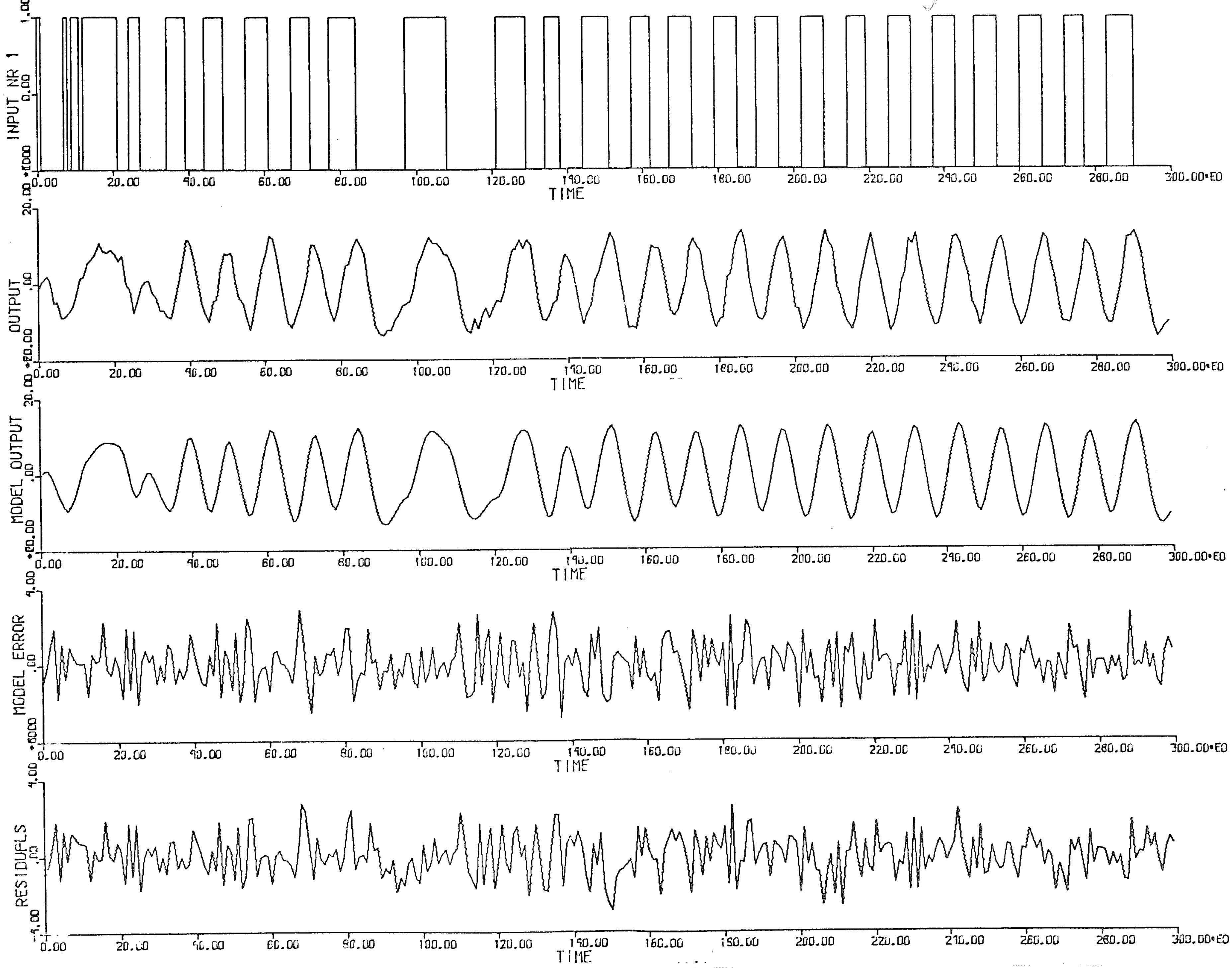


Figure 27



Parameters	True values	Parameter estimations and standard deviations for PRBS input		for D-optimum input	
a_1	-0.5	-0.5263	0.06819	-0.5264	0.05824
b_1	0.5	0.4763	0.04642	0.4658	0.04139
c_1	-0.2	-0.2666	0.08651	-0.2641	0.07679
λ	1.0	1.0457		1.0452	
V	250.0	273.3662		273.1278	
$\lvert \underline{\underline{K}}_N \rvert$		$8.2370\ 10^{-6}$		$4.8178\ 10^{-6}$	
$tr(\underline{\underline{K}}_N)$		$6.2231\ 10^{-3}$		$4.6728\ 10^{-3}$	

Table V. N=500

Parameters	True values	Parameter estimations and standard deviations for PRBS input		for D-optimum input	
a_1	-0.5	-0.5061	0.06257	-0.5004	0.05437
b_1	0.5	0.4744	0.04728	0.5279	0.04621
c_1	0.0	-0.0445	0.07666	-0.0370	0.06951
λ	1.0	1.04573		1.04568	
V	250.0	273.3911		273.3626	
$\lvert \underline{\underline{K}}_N \rvert$		$7.2683\ 10^{-6}$		$5.2792\ 10^{-6}$	
$tr(\underline{\underline{K}}_N)$		$5.6243\ 10^{-3}$		$4.6563\ 10^{-3}$	

Table VI. N=500

Para-meters	True values	Parameter estimations and standard deviations			
		for PRBS input		for D-optimum input	
a_1	-0.5	-0.5282	0.05502	-0.4992	0.04618
b_1	0.5	0.4779	0.04515	0.5190	0.03604
c_1	-0.5	-0.5611	0.06350	-0.5321	0.05816
λ	1.0	1.04566		1.04567	
V	250.0	273.35060		273.35380	
$\lvert \underline{\underline{K}}_N \rvert$		$4.0412\ 10^{-6}$		$1.8840\ 10^{-6}$	
$tr(\underline{\underline{K}}_N)$		$4.6330\ 10^{-3}$		$3.1384\ 10^{-3}$	

Table VII. N=500

Para-meters	True values	Parameter estimations and standard deviations			
		for PRBS input		for D-optimum input	
a_1	-1.5	-1.4964	0.005806	-1.5065	0.002919
a_2	0.7	0.6982	0.004177	0.7064	0.002087
b_1	1.0	1.0023	0.036880	0.9720	0.019300
b_2	0.5	0.5114	0.05160	0.5163	0.02714
c_1	-1.8	-1.7981	0.01244	-1.8125	0.008687
c_2	0.9	0.8954	0.01141	0.9111	0.007558
λ	1.0	1.0454		1.0346	
V	250.0	273.2299		267.6178	
$\lvert \underline{\underline{K}}_N \rvert$		$2.5711 \; 10^{-18}$		$4.3549 \; 10^{-21}$	
$tr(\underline{\underline{K}}_N)$		$3.7276 \; 10^{-3}$		$1.0480 \; 10^{-3}$	

Table VIII. N=500

Example 6.2.5.

The system equation :

$$(1-1.5z^{-1}+0.7z^{-2})y(t) = (1.0z^{-1}+0.5z^{-2})u(t) + \lambda(1-1.5z^{-1}+0.7z^{-2})e(t) \qquad (6.2.5)$$

The results of ML identification can be seen in Table IX.
In the case of identification of system (6.2.5) the time functions of input, output, model output,model error and residuals are shown on Figs.28.,29. for PRBS and D-optimum input signal,respectively. It can be established from Figs.26.,27.,28. and 29. that the D-optimum input series contain mostly lower frequency components whose physical interpretation can be given very easily. Namely the applied noise model corresponds to an upper pass filter;the process, however, to a lower pass filter. From this it follows that the maximum signal/noise ratio can be reached rather in the case of lower frequencies than higher frequencies.

Example 6.2.6.

The simulated system equation is:

$$(1-1.5z^{-1}+0.7z^{-2})y(t) = (1.0z^{-1}+0.5z^{-2})u(t) + \lambda(1+1.5z^{-1}+0.7z^{-2})e(t) \qquad (6.2.6)$$

The ML estimates are in Table X.

Figure 28

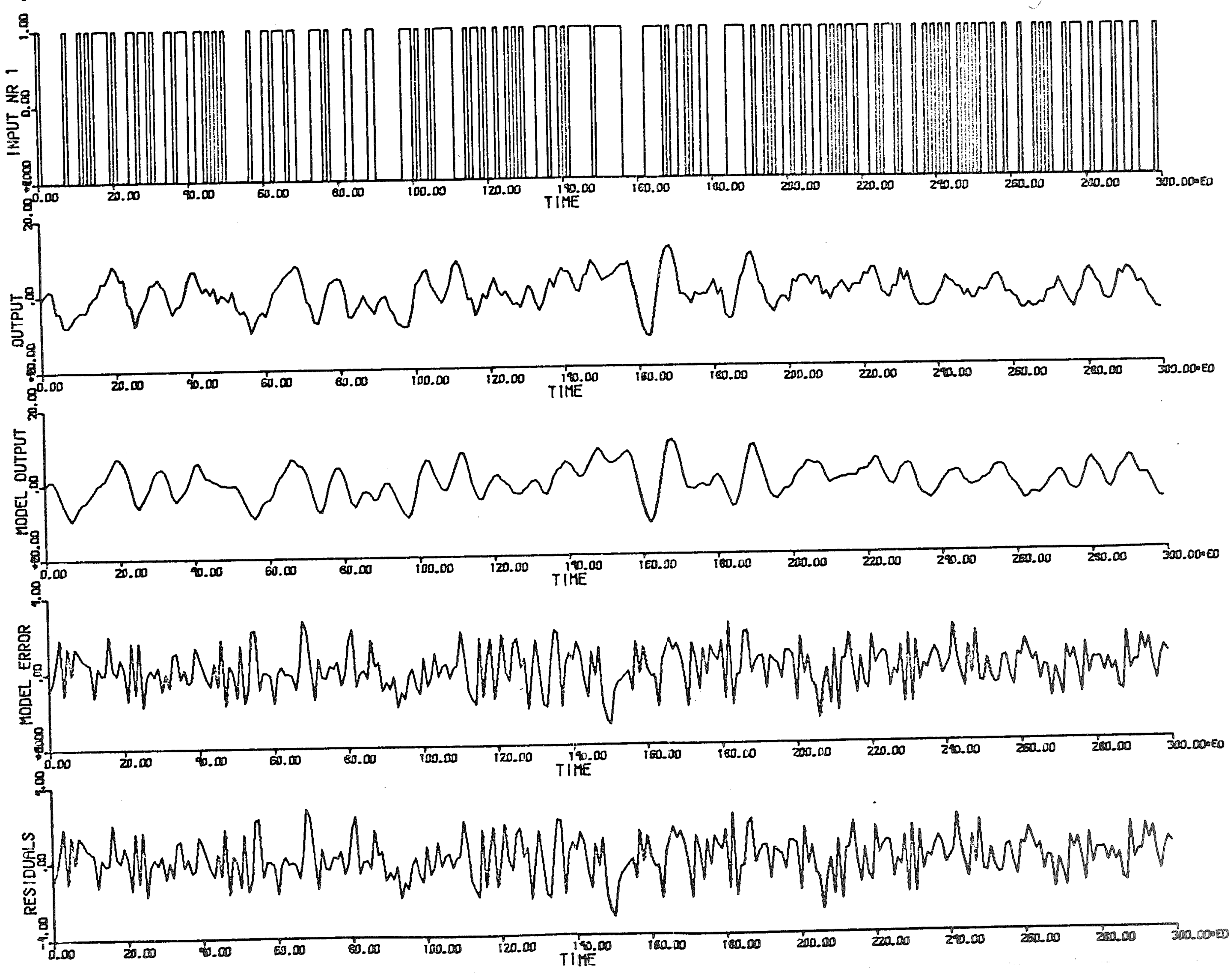
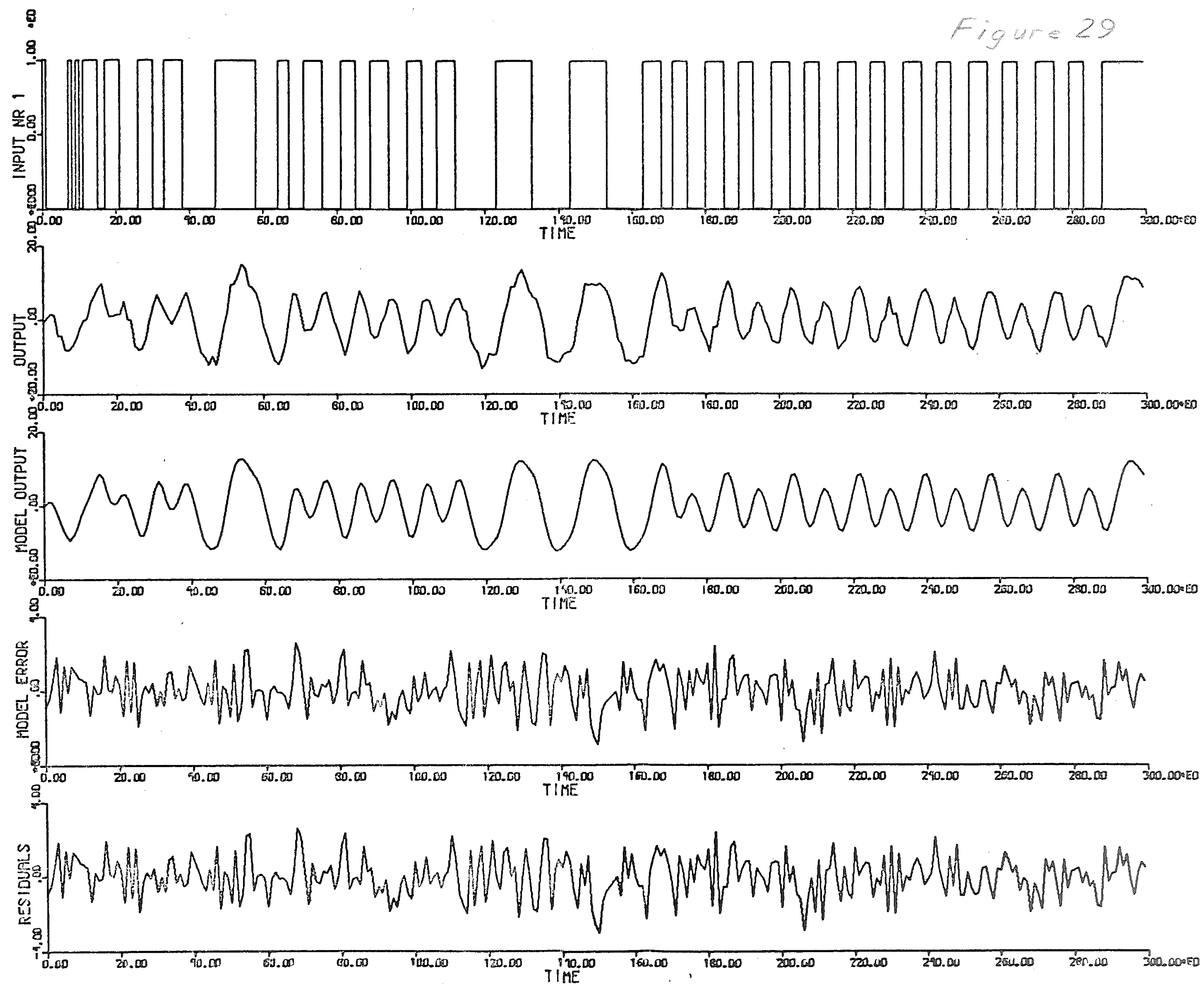


Figure 29



INPUT NR 1
1.00
0.00
OUTPUT
MODEL OUTPUT
MODEL ERROR
RESIDUALS
TIME
0.00
20.00
40.00
60.00
80.00
100.00
120.00
140.00
160.00
180.00
200.00
220.00
240.00
260.00
280.00
300.00+E0

Parameters	True values	Parameter estimations and standard deviations for PRBS input		for D-optimum input	
a_1	-1.5	-1.4948	0.008332	-1.4998	0.005283
a_2	0.7	0.6980	0.006449	0.7016	0.004234
b_1	1.0	0.9756	0.04334	0.9813	0.02574
b_2	0.5	0.5509	0.05843	0.5255	0.03512
c_1	-1.5	-1.5229	0.02471	-1.5280	0.02829
c_2	0.7	0.7282	0.02508	0.7279	0.02836
λ	1.0	1.0442		1.0441	
V	250.0	272.5700		272.5540	
$\lvert \underline{K}_N \rvert$		$6.1251\ 10^{-17}$		$1.4861\ 10^{-18}$	
$tr(\underline{K}_N)$		$4.9605\ 10^{-3}$		$1.1714\ 10^{-3}$	

Table IX. N=500

Para-meters	True values	Parameter estimations and standard deviations for PRBS input		for D-optimum input	
a_1	-1.5	-1.4930	0.03428	-1.4975	0.02742
a_2	0.7	0.6925	0.03392	0.6983	0.02723
b_1	1.0	0.9690	0.03545	1.0244	0.02328
b_2	0.5	0.4854	0.03333	0.4986	0.02139
c_1	1.5	1.4694	0.04172	1.4862	0.05523
c_2	0.7	0.6816	0.04248	0.7012	0.05722
λ	1.0	1.0439		1.0391	
V	250.0	272.4219		269.9331	
$\lvert \underline{\underline{K}}_N \rvert$		$4.8995 \; 10^{-14}$		$3.9782 \; 10^{-15}$	
$tr(\underline{\underline{K}}_N)$		$4.3068 \; 10^{-3}$		$2.3096 \; 10^{-3}$	

<u>Table X.</u> N=500

Example 6.2.7.

The simulated process:

$$(1-1.5z^{-1}+0.7z^{-2})y(t) = (1.0z^{-1}+0.5z^{-2})u(t) + \lambda(1-1.0z^{-1}+0.2z^{-2})e(t) \tag{6.2.7}$$

The ML estimates are presented in Table XI.

It can be established from the simulation results that the CRAMER-RAO lower bound can be more approached by the optimum input signal series than by PRBS.

It would require further investigations to determine the sensitivity of the input signal synthesis to the apriori estimates (known from the previous estimation) of the parameters.

Para-meters	True values	Parameter estimations and standard deviations			
		for PRBS input		for D-optimum input	
a_1	-1.5	-1.5019	0.01312	-1.5032	0.01235
a_2	0.7	0.7055	0.01119	0.6914	0.01022
b_1	1.0	0.9720	0.04795	0.9984	0.04091
b_2	0.5	0.5399	0.06229	0.4947	0.05401
c_1	-1.0	-1.0409	0.03874	-1.0454	0.04260
c_2	0.2	0.2308	0.03922	0.2175	0.04210
λ	1.0	1.0445		1.0430	
V	250.0	272.7218		271.9616	
$\lvert \underline{\underline{K}}_N \rvert$		$2.1924 \; 10^{-15}$		$8.5434 \; 10^{-16}$	
$tr(\underline{\underline{K}}_N)$		$5.9384 \; 10^{-3}$		$4.4561 \; 10^{-3}$	

Table XI. N=500

7. SUMMARY AND CONCLUSIONS

The simulation results have proved the truth of the theoretical investigations and the fact that it is worth dealing with the synthesis of optimum input signal. The maximization of the determinant of the information matrix has proved to be a good criterion and the rule of the synthesis corresponds to this strategy is simple.

The on-line input signal synthesis has rather theoretical importance and its application is the question of future.

The off-line input signal synthesis, however, is worth using on the following manner. Assume that we have performed an ML estimation with non-optimal input signals then generate a D-optimum series by these estimates. After calculating the spectra of the initial and optimal inputs we compare them. If they are nearly the same then the original input is suitable, if they not then the original one has to be modified according to the optimal spectrum or the optimum input series is to be applied. It can be said in the basis of simulation results that such procedure can improve the accuracy of the estimates significantly.

The noise model used to the investigation of the off-line input signal synthesis was upper pass filter and the process, however, was lower pass one, therefore rather the lower frequency input serie was optimal and effective for the synthesis.

In the case of on-line synthesis both the noise model and process model were upper pass filter (derivative type) so there was not considerable difference between the spectra of inputs and the accuracy of the estimates did not improved in so high degree.

In the future the following questions are to be analysed:

1. We have to investigate the applicability of the "one stage" and "N-stage" control used at the optimal control of discrete-time systems for the criteria of input synthesis and their similarities to the algorithms of the synthesis.
2. We have to deal with the question of "persistently exiting" feature in connection with the optimal input series.
3. The identifiability question of parameters must be examined in the case of on-line input signal synthesis.
4. We have to analyse the sensitivity of optimal input to the apriori parameter estimates.
5. We have to study what possibilities are to develop simple algorithms of input signal synthesis for other criteria (A- ,G-optimality).

Inspite of these remaining problems we hope that this report promotes the efforts directed to the improving of estimates and gives simple,application-oriented methods for the input signal synthesis in the case of linear discrete-time dynamic systems.

8. ACKNOWLEDGEMENTS

I would like to thank Professor Frigyes Csaki for making possible my study-tour and initiating this problem.
I am also very grateful Professor Karl Johan Åström who provided opportunity to join in the research work of Division of Automatic Control, Lund, to work on the computer UNIVAC 1108 of the University and to use the program-package worked out at the Department.
I wish to express my gratitude to I. Gustavsson for valuable suggestions and discussions.

9. REFERENCES

[1] F. Csaki : "Modern Control Theory", Academic Press, Budapest, 1972.

[2] K.J. Åström : "Introduction to Stochastic Control Theory", Academic Press, New York and London, 1970.

[3] K.J. Åström : "On the Achievable Accuracy in Identification Problems", IFAC, Prague, 1967.

[4] K.J. Åström and T.Bohlin : "Numerical Identification of Linear Dynamic Systems from Normal Operating Records", IFAC, Teddington, 1965.

[5] K.J. Åström and P.Eykhoff : "System Identification - A Survey", Automatica, Vol.7, 1971, 123-162.

[6] I.Gustavsson : "Choice of Sampling Interval for Parametric Identification", Report 7103, Division of Automatic Control, Lund Institute of Technology, 1971.

[7] I. Gustavsson : "Parametric Identification of Multiple Input, Single Output Linear Dynamic Systems", Report 6907, Division of Automatic Control, Lund Institute of Technology, 1969.

[8] J. Kiefer and J. Wolfowitz : "Optimum Designs in Regression Problems", Annals of Math. Stat., Vol.30, 1959, 271-294.

[9] M. Stone : "Application of a Measure of Information to the Design and Comparison of Regression Experiments", Annals of Math. Stat., Vol.30, 1959, 55-70.

[10] S.N. Sokolov : "Continuous Design of Regression Experiments", Theory of Probability,Vol.8,1963,95-101,318-324.

[11] S. Kullback : "Information Theory and Statistics",John Wiley and Sons,New York,1959.

[12] M. Aoki and R.M. Staley : "On Input Signal Synthesis in Parameter Identification",Automatica,Vol.6,1970, 431-440.

[13] M. Aoki : "Optimization of Stochastic System",Academic Press, New York and London,1967.

[14] V.S. Levadi : "Design of Input Signals for Parameter Estimation",IEEE Trans.Aut.Control,AE-11,Vol.2,1966.

[15] R.M. Gagliardi : "Input Selections for Parameter Identification in Discrete Systems",IEEE Trans.Aut.Control. AC-12,Vol.5,1967.

[16] A. Rault,R. Poliquen and J. Richalet : "Sensibilizing input and Identification",IFAC,Warsaw,1969.

[17] K. Inone,K. Ogino and Y. Sawaragi : "Sensitivity Synthesis of Optimal Input for Parameter Identification", IFAC,Prague,1970.

[18] M. Aoki andP.C. Yue : "On Certain Convergence Questions in System Identification",Siam Journal on Control,Vol.2, 1970,239-256.

[19] C.P. Rao : "Linear Statistical Inference and its Applications", John Wiley,New York,1967.

[20] L.K. Keviczky : "Design of Regression Experiments",Doctor Thesis, Technical University,Budapest,1970.

[21] L.K. Keviczky and Cs. Banyasz : "Optimal Identification by Simulation of the Information Obtained from Processes",Summer Simulation Conference,San Diego, 1972.

A P P E N D I X

APPENDIX A.

Assuming that $\underline{\varepsilon}$ has normal distribution the information matrix can be determined easily. Then taking into account (3.2) and (3.4) the probability density function is:

$$L(\underline{y}|\underline{p}) = k_1 \exp\left\{-\frac{1}{2}(\underline{y}-\underline{x})^T \underline{\underline{W}} (\underline{y}-\underline{x})\right\} \qquad (A.1)$$

Thus

$$\ln L(\underline{y}|\underline{p}) = k_2 - \frac{1}{2}(\underline{y}-\underline{x})^T \underline{\underline{W}} (\underline{y}-\underline{x}) \qquad (A.2)$$

Hence

$$\frac{d \ln L(\underline{y}|\underline{p})}{d\underline{p}} = \frac{d\underline{x}^T}{d\underline{p}} \frac{d \ln L(\underline{y}|\underline{p})}{d\underline{x}} = \frac{d\underline{x}^T}{d\underline{p}} \underline{\underline{W}} (\underline{y}-\underline{x}) =$$

$$= \frac{d\underline{x}^T}{d\underline{p}} \underline{\underline{W}} \underline{\varepsilon} \qquad (A.3)$$

Taking the definition (3.3) of $\underline{\underline{M}}$ and (3.4)

$$\underline{\underline{M}} = \int L(\underline{y}|\underline{p}) \left[\frac{d\underline{x}^T}{d\underline{p}} \underline{\underline{W}} \underline{\varepsilon}\right]\left[\frac{d\underline{x}^T}{d\underline{p}} \underline{\underline{W}} \underline{\varepsilon}\right]^T d\underline{y} =$$

$$= \frac{d\underline{x}^T}{d\underline{p}} \underline{\underline{W}} E\left\{\underline{\varepsilon}\,\underline{\varepsilon}^T\right\} \underline{\underline{W}}^T \frac{d\underline{x}}{d\underline{p}^T} = \frac{d\underline{x}^T}{d\underline{p}} \underline{\underline{W}} \frac{d\underline{x}}{d\underline{p}^T} \qquad (A.4)$$

where it has been used that $\underline{\underline{W}}$ is symmetrical.

APPENDIX B.

Assuming that the parameter estimate $\hat{\underline{p}}_N = \hat{\underline{p}}|\underline{y}_N$ has normal distribution with $\underline{p}$ mean value and $\underline{\underline{K}}$ covariance matrix then the probability density function is:

$$L(\hat{\underline{p}}_N) = L(\hat{\underline{p}}|\underline{y}_N) = \sqrt{\frac{|\underline{\underline{K}}_N^{-1}|}{(2\pi)^k}} \exp\left\{-\frac{1}{2}(\hat{\underline{p}} - \underline{p})^T \underline{\underline{K}}_N^{-1} (\hat{\underline{p}} - \underline{p})\right\} \tag{B.1}$$

where k is the dimension of $\hat{\underline{p}}$ and

$$E\left\{(\hat{\underline{p}}_N - \underline{p})(\hat{\underline{p}}_N - \underline{p})^T\right\} = \underline{\underline{K}}_N \tag{B.2}$$

The logarithm of (B.2)

$$\log\left\{L(\hat{\underline{p}}_N)\right\} = \frac{1}{2}\log|\underline{\underline{K}}_N^{-1}| - \frac{k}{2}\log(2\pi) - \frac{1}{2}(\hat{\underline{p}} - \underline{p})^T \underline{\underline{K}}_N^{-1} (\hat{\underline{p}} - \underline{p}) \tag{B.3}$$

Let us determine the expectation value of (B.3) according to the definition (3.18). The first two terms of the right side are constant and the expectation value of the third term:

$$E\left\{(\hat{\underline{p}}_N - \underline{p})^T \underline{\underline{K}}_N^{-1} (\hat{\underline{p}}_N - \underline{p})\right\} = E\left\{tr\left[\underline{\underline{K}}_N^{-1}(\hat{\underline{p}}_N - \underline{p})(\hat{\underline{p}}_N - \underline{p})^T\right]\right\} =$$

$$= tr\left[\underline{\underline{K}}_N^{-1} E\left\{(\hat{\underline{p}}_N - \underline{p})(\hat{\underline{p}}_N - \underline{p})^T\right\}\right] = tr\left[\underline{\underline{K}}_N^{-1}\underline{\underline{K}}_N\right] = k \tag{B.4}$$

That is on the basis of (3.18),(B.3) and (B.4)

$$I\left\{L(\hat{\underline{p}}|\underline{y}_N)\right\} = \frac{1}{2}\log|\underline{\underline{K}}_N^{-1}| - k\,\frac{1 + \log(2\pi)}{2} \tag{B.5}$$

In according to this

$$\Delta I = I\left\{L(\hat{\underline{p}}|\underline{y}_{N+1})\right\} - I\left\{L(\hat{\underline{p}}|\underline{y}_N)\right\} = \frac{1}{2}\log\frac{|\underline{\underline{K}}_{N+1}^{-1}|}{|\underline{\underline{K}}_N^{-1}|} \tag{B.6}$$

APPENDIX C.

Taking into account the vector equation (5.2) for N samplings it can be written that

$$\underline{y} = \underline{x} + \underline{\varepsilon} \tag{C.1}$$

where

$$\underline{\varepsilon} = \lambda \underline{\underline{A}}^{-1} \underline{\underline{C}}\ \underline{e} \tag{C.2}$$

and

$$\underline{x} = \underline{\underline{A}}^{-1} \underline{\underline{B}}\ \underline{u} \tag{C.3}$$

So thus

$$E\{\underline{\varepsilon}\ \underline{\varepsilon}^T\} = \lambda^2 \underline{\underline{A}}^{-1} \underline{\underline{C}} (\underline{\underline{A}}^{-1} \underline{\underline{C}})^T = \underline{\underline{W}}^{-1} \tag{C.4}$$

Let us determine the JACOBIAN-matrix $\frac{d\underline{x}^T}{d\underline{p}}$ for the computation of information matrix according to (3.5).

The derivatives with respect to the components are:

$$\frac{\partial \underline{x}}{\partial a_i} = \frac{\partial \underline{\underline{A}}^{-1}}{\partial a_i} \underline{\underline{B}}\ \underline{u} = - \underline{\underline{A}}^{-1} \frac{\partial \underline{\underline{A}}}{\partial a_i} \underline{\underline{A}}^{-1} \underline{\underline{B}}\ \underline{u} = - \underline{\underline{A}}^{-1} \underline{\underline{S}}^i\ \underline{x} \tag{C.5}$$

and

$$\frac{\partial \underline{x}}{\partial b_i} = \underline{\underline{A}}^{-1} \frac{\partial \underline{\underline{B}}}{\partial b_i} \underline{u} = \underline{\underline{A}}^{-1} \underline{\underline{S}}^i\ \underline{u} \tag{C.6}$$

where the definition of $\underline{\underline{A}}$ and $\underline{\underline{B}}$ according to the (5.3) and (5.4) has been applied.

So

$$\frac{d\underline{x}^T}{d\underline{p}} = \begin{bmatrix} \underline{u}^T (\underline{\underline{S}}^T)^o \\ \underline{u}^T \vdots (\underline{\underline{S}}^T)^m \\ \underline{x}^T (\underline{\underline{S}}^T) \\ \vdots \\ \underline{x}^T (\underline{\underline{S}}^T)^n \end{bmatrix} (\underline{\underline{A}}^{-1})^T = \underline{\underline{H}}^T (\underline{u},\underline{x}) (\underline{\underline{A}}^{-1})^T \tag{C.7}$$

In according to this the information matrix is:

$$\underline{\underline{M}} = \underline{\underline{H}}^T (\underline{u},\underline{x}) (\underline{\underline{A}}^{-1})^T [\lambda^2 \underline{\underline{A}}^{-1} \underline{\underline{C}} (\underline{\underline{A}}^{-1}\underline{\underline{C}})^T]^{-1} \underline{\underline{H}}(\underline{u},\underline{x}) \underline{\underline{A}}^{-1} =$$

$$= \frac{1}{\lambda^2} \underline{\underline{H}}^T(\underline{u},\underline{x}) (\underline{\underline{C}}^{-1})^T \underline{\underline{C}}^{-1} \underline{\underline{H}}(\underline{u},\underline{x}) \qquad (C.8)$$

The multiplying with $\underline{\underline{C}}^{-1}$ means actually filtering, too,so it can be written that

$$\underline{\underline{M}} = \frac{1}{\lambda^2} \underline{\underline{H}}^T(\underline{u}^F,\underline{x}^F) \underline{\underline{H}}(\underline{u}^F,\underline{x}^F) = \frac{1}{\lambda^2} \sum_{t=1}^{N} \underline{f}(u^F,x^F,t) \underline{f}^T(u^F,x^F,t) \qquad (C.9)$$

where

$$\underline{u}^F = \underline{\underline{C}}^{-1} \underline{u} \quad \text{and} \quad \underline{x}^F = \underline{\underline{C}}^{-1} \underline{x} \qquad (C.10)$$

(see (3.14)) i.e. the corresponding scalar equations are:

$$u^F(t) = \frac{1}{C(z^{-1})} u(t) \quad \text{and} \quad x^F(t) = \frac{1}{C(z^{-1})} x(t) \qquad (C.11)$$

The $\underline{f}(u^F,x^F,t)$ is the following:

$$\underline{f}(u^F,x^F,t) = [u^F(t),\ldots,u^F(t-m),-x^F(t-1),\ldots,-x^F(t-n)]^T \qquad (C.12)$$

The information matrix can be obtained easily from (C.9) in the case of different noise structures.

APPENDIX D.

Let the system equation be:

$$(1+az^{-1})\ y(t) = b_o\ u(t) - \lambda e(t) \tag{D.1}$$

The information matrix is [3,6]

$$\underline{\underline{M}} = \frac{N}{\lambda^2} \begin{bmatrix} E\{(\frac{\partial \varepsilon}{\partial a})^2\} & E\{\frac{\partial \varepsilon}{\partial a}\ \frac{\partial \varepsilon}{\partial b_o}\} \\ E\{\frac{\partial \varepsilon}{\partial b_o}\ \frac{\partial \varepsilon}{\partial a}\} & E\{(\frac{\partial \varepsilon}{\partial b_o})^2\} \end{bmatrix} \tag{D.2}$$

where

$$\varepsilon(t) = (1 + az^{-1})\ y(t) - b_o\ u(t) \tag{D.3}$$

The partial derivatives are:

$$\frac{\partial \varepsilon(t)}{\partial a} = z^{-1}\ y(t) = \frac{b_o\ z^{-1}}{1 + az^{-1}}\ u(t) + \frac{\lambda z^{-1}}{1 + az^{-1}} e(t) \tag{D.4}$$

$$\frac{\partial \varepsilon(t)}{\partial b_o} = -u(t) \tag{D.5}$$

Assuming that u(t) is white noise with variance one and independent of e(t):

$$E\{(\frac{\partial \varepsilon}{\partial a})^2\} = \frac{1}{2\pi i}\oint \frac{b_o z^{-1}}{1+az^{-1}}\ \frac{b_o z}{1+az}\ \frac{dz}{z} + \frac{1}{2\pi i}\oint \frac{\lambda z^{-1}}{1+az^{-1}}\ \frac{\lambda z}{1+az}\ \frac{dz}{z} = \frac{b_o^2 + \lambda^2}{1 - a^2} \tag{D.6}$$

$$E\left\{\frac{\partial \varepsilon}{\partial a}\frac{\partial \varepsilon}{\partial b}\right\} = \frac{-1}{2\pi i}\oint \frac{b_o z^{-1}}{1+az^{-1}}\, 1\, \frac{dz}{z} = 0 \qquad (D.7)$$

$$E\left\{\left(\frac{\partial \varepsilon}{\partial b_o}\right)^2\right\} = \frac{1}{2\pi i}\oint 1\, \frac{dz}{z} = 1 \qquad (D.8)$$

The inverse of information matrix

$$\underline{\underline{M}}^{-1} = \frac{\lambda^2}{N}\begin{bmatrix} \frac{b_o^2 + \lambda^2}{1-a^2} & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \frac{\lambda^2}{N}\begin{bmatrix} \frac{1-a^2}{b_o^2+\lambda^2} & 0 \\ 0 & 1 \end{bmatrix} \qquad (D.9)$$

Hence

$$|\underline{\underline{M}}_N^{-1}| = \frac{\lambda^4}{N^2}\,\frac{1-a^2}{b_o^2+\lambda^2} \qquad (D.10)$$

and

$$tr(\underline{\underline{M}}_N^{-1}) = \frac{\lambda^2}{N}\,\frac{1-a^2+b_o^2+\lambda^2}{b_o^2+\lambda^2} \qquad (D.11)$$

APPENDIX E.

Let the system equation be :

$$(1+az^{-1})y(t) = bz^{-1}\,u(t) + \lambda(1+cz^{-1})\,e(t) \tag{E.1}$$

The information matrix is similar to (D.2).

Writing down $\varepsilon(t)$

$$\varepsilon(t) = \frac{1+az^{-1}}{1+cz^{-1}}\,y(t) - \frac{bz^{-1}}{1+cz^{-1}}\,u(t) \tag{E.2}$$

the derivatives are the followings:

$$\frac{\partial\varepsilon(t)}{\partial a} = \frac{z^{-1}}{1+cz^{-1}}\,y(t) = \frac{bz^{-2}}{(1+cz^{-1})(1+az^{-1})}u(t) + \lambda\frac{z^{-1}}{(1+az^{-1})}e(t) \tag{E.3}$$

$$\frac{\partial\varepsilon(t)}{\partial b} = -\frac{z^{-1}}{1+cz^{-1}}\,u(t) \tag{E.4}$$

Assuming that u(t) is white noise with zero mean and variance one and independent of e(t) we get that

$$E\left\{\left(\frac{\partial\varepsilon}{\partial a}\right)^2\right\} = \frac{1}{2\pi i}\oint \frac{bz^{-2}}{(1+cz^{-1})(1+az^{-1})}\;\frac{bz^{2}}{(1+cz)(1+az)}\;\frac{dz}{z} +$$

$$+\frac{1}{2\pi i}\oint \frac{\lambda\,z^{-1}}{(1+az^{-1})}\;\frac{\lambda\,z}{(1+az)}\;\frac{dz}{z} =$$

$$= \frac{b^2(1+ac) + \lambda^2(1-c^2)(1-ac)}{(1-c^2)(1-a^2)\,(1-ac)} \tag{E.5}$$

$$E\left\{\left(\frac{\partial\varepsilon}{\partial b}\right)^2\right\} = \frac{1}{2\pi i}\oint \frac{z^{-1}}{(1+cz^{-1})}\,\frac{z}{(1+cz)}\,\frac{dz}{z} = \frac{1}{1-c^2} \qquad (E.6)$$

$$E\left\{\frac{\partial\varepsilon}{\partial a}\,\frac{\partial\varepsilon}{\partial b}\right\} = \frac{-1}{2\pi i}\oint \frac{z^{-1}}{(1+cz^{-1})}\,\frac{bz^2}{(1+cz)(1+az)}\,\frac{dz}{z} = \frac{bc}{(1-ac)(1-c^2)} \qquad (E.7)$$

The inverse of the information matrix is :

$$\underline{\underline{M}}^{-1} = \frac{\lambda^2}{N}\begin{bmatrix} \dfrac{b^2(1+ac)+\lambda^2(1-c^2)(1-ac)}{(1-c^2)(1-a^2)\,(1-ac)} & \dfrac{bc}{(1-c^2)(1-ac)} \\ \dfrac{bc}{(1-c^2)(1-ac)} & \dfrac{1}{(1-c^2)} \end{bmatrix}^{-1} =$$

$$= \frac{\lambda^2(1-ac)}{N\left[b^2+\lambda^2(1-ac)^2\right]}\begin{bmatrix} (1-a^2)(1-ac) & -bc(1-a^2) \\ -bc(1-a^2) & b^2(1+ac)+\lambda^2(1-c^2)(1-ac) \end{bmatrix} \qquad (E.8)$$

Hence

$$\left|\underline{\underline{M}}_N^{-1}\right| = \frac{\lambda^4(1-ac)^2(1-a^2)(1-c^2)}{N^2\left[b^2+\lambda^2(1-ac)^2\right]} \qquad (E.9)$$

and

$$\mathrm{tr}(\underline{\underline{M}}_N^{-1}) = \frac{\lambda^2(1-ac)\left\{(1-a^2)(1-ac)+b^2(1+ac)+\lambda^2(1-c^2)(1-ac)\right\}}{N\left[b^2+\lambda^2(1-ac)^2\right]} \qquad (E.10)$$